UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO" – CAMPUS BOTUCATU
INSTITUTO DE BIOCIÊNCIAS

Histórico da ocupação das chapadas brasileiras por espécies do gênero *Rauvolfia* L. (Apocynaceae) e suas implicações biogeográficas

TESE DE DOUTORADO

JOÃO DE DEUS VIDAL JÚNIOR

BOTUCATU – SP
2018

Histórico da ocupação das chapadas brasileiras por espécies do gênero *Rauvolfia* L. (Apocynaceae) e suas implicações biogeográficas

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO -CRB 8/7500

Vidal, João de Deus.
Histórico da ocupação das chapadas brasileiras por espécies do gênero *Rauvolfia L.* (Apocynaceae) e suas implicações biogeográficas / João de Deus Vidal. - Botucatu, 2018

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Ingrid Koch
Coorientador: Samantha Koehler
Capes: 20304005

1. Microssatélites (Genética). 2. Genética de populações. 3. Cerrados. 4. Marcadores genéticos. 5. Zonas biológicas.

Palavras-chave: Cerrado; Genética de Populações; Microssatélites; Modelagem de nicho.

JOÃO DE DEUS VIDAL JÚNIOR

Histórico da ocupação das chapadas brasileiras por espécies do gênero *Rauvolfia* L. (Apocynaceae) e suas implicações biogeográficas

Tese de doutorado apresentada ao Instituto de Biociências, campus de Botucatu, UNESP, para obtenção do título de Doutor no Programa de Pós-Graduação em Ciências Biológicas (Botânica). Área de concentração: Sistemática e Ecologia Vegetal.

Profa. Dra. Ingrid Koch
Orientadora

Prof. Dra. Samantha Koehler
Co-orientadora

BOTUCATU – SP

2018

"(…) *os seres humanos não nascem para sempre*

no dia em que as mães os dão à luz, e sim que a vida os obriga outra vez

e muitas vezes a parirem a si mesmos."

Gabriel García Márquez

Para Denise, João, Dona Maria, Zinho e Ditinha,

que me ensinaram o caminho das flores.

AGRADECIMENTOS

Inicialmente, gostaria de agradecer às agências de fomento CAPES e FAPESP, que em diferentes oportunidades financiaram custos deste trabalho. À agência CAPES, um agradecimento especial pela bolsa recebida.

Meus sinceros agradecimentos também aos funcionários e docentes do Departamento de Biologia Vegetal da UNICAMP e do Departamento de Botânica da UNESP Botucatu, que permitiram que eu tivesse condições de desenvolver as atividades necessárias a minha pós-graduação.

Este trabalho só foi possível graças ao grande empenho das professoras Ingrid Koch, Samantha Koehler e Anete Pereira. Portanto, meus maiores agradecimentos vão para elas, que depositaram tempo, esforço, recursos e paciência em me auxiliar na elaboração desta tese. Um agradecimento especial à professora Ingrid, que me acompanha na vida acadêmica desde a minha iniciação científica, há quase 10 anos atrás. Sem seu apoio e a sua capacidade de gerenciar os desafios de uma forma tão competente (e, ao mesmo tempo, amável) estes anos teriam sido, sem dúvida, muito mais difíceis. Agradeço também aos membros da banca, prof. Fábio Pinheiro, profa. Maria Fernanda Calió, prof. Alessandro Alves e profa. Maria Zucchi, pela disponibilidade de contribuir com esta etapa final do trabalho.

Agradeço aos meus pais pela estrutura afetiva, acolhimento e sacrifícios financeiros que fizeram desde que eu ingressei na escola. Sei que vocês abriram mão de muitas coisas para que eu e meus irmãos pudéssemos chegar onde chegamos, e a minha forma de gratidão é ter feito meu melhor para aproveitar cada oportunidade que o suor de vocês criou. Este tese só

foi possível graças a todas as vezes que vocês estudaram tabuada comigo, mesmo cansados do trabalho e da dura rotina de criar quatro filhos. Um agradecimento especial ao meu pai, que foi meu companheiro de coletas e agente financiador nas duas expedições ao Mato Grosso. Agradeço aos meus avós Zinho e Maria, que nos deixaram no ano passado, mas cuja paixão pelas plantas desde muito cedo me cativou e fez germinar em mim o amor pela botânica. Sempre me lembro do jardim cheio de orquídeas, da cerca-viva de tumbérgias-azuis, do cheiro de dama da noite na casa da vó Ditinha, e das hortas repletas de boldo, hortelã e outras ervas que a vó Maria cultivava. Carrego vocês pra sempre dentro do coração.

Agradeço também aos colegas e técnicos de laboratório do LAGM: Yohans, Aline, Danilo, Bia, Carla, Rebecca, Liana, Fernanda, Stephanie, Fábio, Luciano, Lívia, André e a todos os outros com quem compartilhei um bolo com café nas tardes destes últimos anos. Vocês são os melhores colegas de laboratório que alguém poderia ter. Muito obrigado pela força após cada PCR falhada, pela paciência após cada placa de acrilamida quebrada e pela ajuda de todos os dias.

Finalmente, agradeço aos grandes amigos que a vida me deu na pós-graduação, tanto em Botucatu quanto em Campinas. André, Lucas, Nicole, Mari, Rãxera, Rita, Fosseteira, entre outros, com quem compartilhei muitas mesas de bar nestes quatro anos. Um agradecimento especial a minha namorada Bruna, que esteve do meu lado nos momentos mais difíceis neste último ano e me apoiou sempre que precisei. Vocês são a razão de eu ter conseguido seguir em frente.

RESUMO

VIDAL Jr., João de Deus. *Histórico da ocupação das chapadas brasileiras por espécies do gênero Rauvolfia L. (Apocynaceae) e suas implicações biogeográficas*. 2018. 131 f. Tese de Doutorado em Ciências Biológicas (Botânica) – Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu, Botucatu, 2017.

Neste estudo, nós desenvolvemos marcadores moleculares e aplicamos abordagens integrativas em filogeografia e ecologia para investigar e descrever a história biogeográfica das chapadas da região oeste do Cerrado brasileiro. Utilizando como modelo duas espécies do gênero *Rauvolfia* L. (Apocynaceae, Rauvolfioideae), *Rauvolfia weddelliana* Müll.Arg. e *Rauvolfia gracilis* I.Koch & Kin.-Gouv., típicas destes ambientes, nós desenvolvemos um conjunto de marcadores nucleares microssatélites e sequenciamos espaçadores de cloroplasto, com os quais estimamos métricas de diversidade genética e estruturação populacional. A partir destes dados, nós reconstruímos relações filogenéticas entre as linhagens, estimamos tempos de divergência e realizamos simulações demográficas e modelos de distribuição, com o intuito de testar possíveis cenários biogeográficos. Em sequência, nós desenvolvemos modelos de distribuição potencial em escala de comunidade e análises espaciais, para estimar o impacto relativo das flutuações climáticas do Pleistoceno sobre o endemismo no Cerrado, quando comparadas a outros processos. Os resultados estão apresentados em três capítulos. No primeiro, intitulado "Development and cross-validation of microsatellite markers for *Rauvolfia weddelliana* Müll.Arg. (Apocynaceae) species complex", nós descrevemos o

desenvolvimento de uma biblioteca de microssatélites específica para *R. weddelliana* e testamos sua transferabilidade para a espécie *R. gracilis*. Nós isolamos dez marcadores transferíveis, com alto conteúdo de polimorfismos que podem agora ser aplicados em estudos genéticos para o grupo, e também demonstramos a diploidia das espécies através do padrão de bandas de cada lócus. No segundo capítulo, intitulado "Phylogeography and paleodistribution of *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I.Koch & Kin.-Gouv. (Apocynaceae)", nós aplicamos estes marcadores em conjunto com espaçadores de cloroplasto para testar, através de análises demográficas e modelos de nicho, os impactos das flutuações climáticas nas populações de *R. weddelliana* e *R. gracilis*. Nossas análises recuperaram duas linhagens com origem anterior ao Pleistoceno, coincidindo com o soerguimento do Planalto Central Brasileiro, que sofreram retração interglacial durante o Pleistoceno. Apesar de nossos modelos de distribuição terem apresentado resultado contrário a retração interglacial, eles recuperaram a mesma divergência norte-sul, também anterior ao último eventos de glaciação. Estes resultados sugerem que as savanas dos planaltos do oeste do Cerrado provavelmente configuram refúgios interglaciais. No terceiro capítulo, nós testamos este padrão, descrevendo a estabilidade relativa de chapadas e vales e modelando as distribuições potenciais de espécies típicas destes ambientes. Este trabalho contribui para a discussão sobre os impactos do clima do Pleistoceno e dos processos geológicos do Mioceno na diversificação do complexo domínio dos Cerrados.

Palavras-chave: Microssatélites. Genética de Populações. Cerrado. Modelagem de nicho.

ABSTRACT

In this study, we developed molecular markers and applied integrative approaches in phylogeography and ecology to investigate and describe the biogeographic history of the plateaus of the west region of the Cerrado in Brazil. Using as models two species of the genus *Rauvolfia* L. (Apocynaceae, Rauvolfioideae), *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I.Koch & Kin.-Gouv., typical from these habits, we developed a set of nuclear microsatellite markers and sequenced chloroplast spacers, which we used to estimate metrics of genetic diversity and population structure. From this data, we reconstructed phylogenetic relationships between lineages, estimated divergence times and conducted demographic simulations and distribution models, intending to test possible biogeographic scenarios. In sequence, we also developed community wise models of potential distribution and conducted spatial analysis to estimate the relative impact of climatic shifts of the Pleistocene on the endemism of the Cerrado, when compared to other processes. Our results are presented in three chapters. In the first, entitled "Development and cross-validation of microsatellite markers for *Rauvolfia weddelliana* Müll.Arg. (Apocynaceae) species complex", we describe the development of a specific microsatellite library to *R. weddelliana* and tested its transferability to the species *R. gracilis*. We isolated ten transferable markers with high polymorphism content that may now be applied to genetics studies on the group, and also demonstrated the diploidy for these species though the patterns of bands for each locus. In the second chapter, entitled "Phylogeography and paleodistribution of *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I.Koch & Kin.-Gouv. (Apocynaceae)", we applied these markers, along with sequences of chloroplast spacers to test, with demographic analysis and

niche models, the impacts of the climatic fluctuations on the populations of *R. weddelliana* and *R. gracilis*. Our analysis recovered two lineages with origin prior to Pleistocene, coinciding with the Brazilian Central Plateau uplift. These lineages underwent interglacial retractions along the Pleistocene. Despite distribution models presented results that are contrary to the interglacial retraction, they recovered the same north-south divergence, also predating the Last Glacial Maximum. These results suggest that the savannas in the plateaus of the western Cerrado are potentially configuring interglacial refugia. In the third chapter, we tested this pattern, by describing the relative stability of plateaus and valleys and modelling the potential distributions of species typical from these environments. This study contributes to the discussions about the impacts of the Pleistocene climate and the geological processes of the Miocene on the diversification the complex Cerrado domain.

Keywords: Microsatellites. Population Genetics. Cerrado. Niche Modelling.

SUMÁRIO

1. Introdução geral

A biogeografia do domínio do Cerrado

O Cerrado é uma vegetação savânica característica das planícies e chapadas do Brasil Central, que ocupa quase 23% da área do país (Ratter *et al*., 1997). É o segundo maior domínio brasileiro, ocupando originalmente mais de 2,000,000 km², sendo superado em extensão apenas pela Floresta Amazônica (Ratter & Dargie, 1992). É caracterizado biologicamente pela notável presença de gramíneas e de uma vegetação esclerófila, xeromórfica, adaptada ao regime de fogo e dotada de sistemas radiculares profundos (e.g. Fiaschi & Pirani, 2009). A paisagem do Cerrado é heterogênea e seu relevo é composto principalmente por planaltos sedimentares compartimentados, intercalados por depressões periféricas (Ab'Saber, 1983). A vegetação savânica típica, ou cerrado *sensu stricto*, está comumente atrelada aos platôs, enquanto matas úmidas estão predominantemente associadas a regiões de vales (Ab'Saber, 1983).

A história biogeográfica desta região é resultado de uma interação complexa entre paisagem, clima e geomorfologia (e.g. Santos *et al*. 2014). Alguns estudos sugerem que as oscilações climáticas do Pleistoceno promoveram ciclos de expansão e coalescência nas vegetações Neotropicais, gerando refúgios dentro das distribuições dos organismos (Vanzolini & Williams 1981; Haffer 1997). Este processo teria resultado no isolamento de linhagens em determinadas regiões (refúgios) durante períodos de condições climáticas desfavoráveis, ocasionando intensa especiação alopátrica em muitos grupos. Alternativamente, outros autores defendem que os eventos geomorfológicos do Terciário, como o soerguimento do Planalto Central brasileiro, teriam impactado de forma mais significativa a diversificação das

linhagens Neotropicais (Prado *et al*. 2012). A expansão das florestas sazonalmente secas e das savanas na região durante os períodos de seca do Pleistoceno é parcialmente suportada por padrões filogeográficos de plantas e animais (Caetano *et al*. 2008, Collevatti *et al*. 2009, Prado *et al*. 2012). Existem, contudo, muitos exemplos de táxons que contrariam este modelo (Santos *et al*. 2004, Vieira *et al*. 2011; Barbosa *et al*. 2015), sugerindo que a ocorrência de retração e fragmentação durante os períodos glaciais foi predominante nestes ambientes. Por conta dessa heterogeneidade de respostas, não há consenso quanto aos impactos das flutuações do clima do Pleistoceno para espécies de Cerrado (e.g. Werneck, 2011; Turchetto-Zolet *et al*. 2013), e sobre como os ciclos glaciais moldaram a diversidade da região e qual a importância relativa destes processos quando comparados a processos geológicos mais antigos.

Abordagens integrativas utilizando marcadores genéticos, simulações computacionais e modelos ecológicos são ferramentas muito uteis para reconstruir a história biogeográfica de populações naturais (Collevatti *et al*. 2012; Bonatelli *et al*. 2014; Peres *et al*. 2015). Dentre os marcadores utilizados para este tipo de investigação, os marcadores microssatélites (em inglês "*short sequence repeats*" – SSR), são pequenas sequências repetitivas de DNA (1 a 6 pares de base), difundidas nos genomas de procariotos e eucariotos (Hamada *et al*. 1982). Este tipo de sequência apresenta elevadas taxas de variabilidade, devido ao fato de a presença de bases repetidas *in tandem* propiciarem uma frequência maior de erros (do tipo "*slippage*") durante a replicação (Levinson & Gutman 1987). Por conta dos elevados níveis de polimorfismo, este tipo de marcador é usualmente aplicado para recuperar informações da história recente de populações (Jarne & Lagoda 1996). Este tipo de marcador tem sido usualmente empregado para investigar efeitos demográficos de flutuações climáticas (Bonatelli *et al*. 2014) e

caracterização da estrutura populacional em populações naturais (Cerón-Souza *et al*. 2012). Modelos de nichos ecológicos, por sua vez, são importantes ferramentas em estudos sobre a distribuição dos organismos em cenários diversos, como climas do passado e do futuro, e tem sido constantemente empregados em estudos filogeográficos (Bonatelli *et al*. 2014; Vieira *et al*. 2015) e macro-ecológicos (Carnaval & Moritz 2008; Werneck *et al*. 2012). Esta metodologia permite o delineamento e o teste de hipóteses sobre a distribuição histórica dos organismos a partir de dados de ocorrência atual, e é uma ferramenta poderosa para auxiliar na compreensão da complexa relação das mudanças climáticas e das origens da biodiversidade.

O gênero *Rauvolfia* L.

O gênero *Rauvolfia* (Apocynaceae, Rauvolfioideae) possui ampla distribuição nos trópicos e três de suas espécies são endêmicas do Cerrado: *Rauvolfia anomala* I.Koch & Rapini, *Rauvolfia weddelliana* Müll.Arg. e *Rauvolfia gracilis* I.Koch & Kin.-Gôuv. (Koch, 2002; Koch *et al*., 2007). Estas espécies compõem um grupo promissor para investigações biogeográficas, uma vez que são facilmente reconhecíveis no campo e porque espécies de *Rauvolfia* são frequentemente diplóides (Raghavan, 1957; Lewis, 1980). Além disso, a distribuição destas espécies no Cerrado engloba diversas áreas de interesse, já reportadas como centros de endemismo ou núcleos de estabilidade climática (Ab'Saber, 1981; Werneck *et al*., 2011), como a Chapada do Guimarães, Chapada dos Parecis, regiões planálticas de Goiás e Mato Grosso do Sul (Rao, 1956, Koch *et al*., 2007, Rapini *et al*., 2010). Dentro deste grupo, *R. weddelliana* possui a distribuição mais ampla, ocorrendo ao longo de praticamente toda a região Centro-Oeste do Brasil e Nordeste do Paraguai, em vegetações campestres e

encostas de montanhas (Rao, 1956). *Rauvolfia anomala* é uma espécie a ser sinonimizada à *R. weddelliana* pois a variação morfológica verificada na espécie vem sendo associada à atuação de fitoplasma durante o desenvolvimento das flores (Vidal, inf. pess.). *Rauvolfia gracilis* possui distribuição mais restrita, ocorrendo em platôs na região norte. *R. weddelliana* e *R. gracilis* são espécies arbustivas, com flores de coloração magenta, permeadas de máculas brancas, morfologicamente variáveis (fig.1).

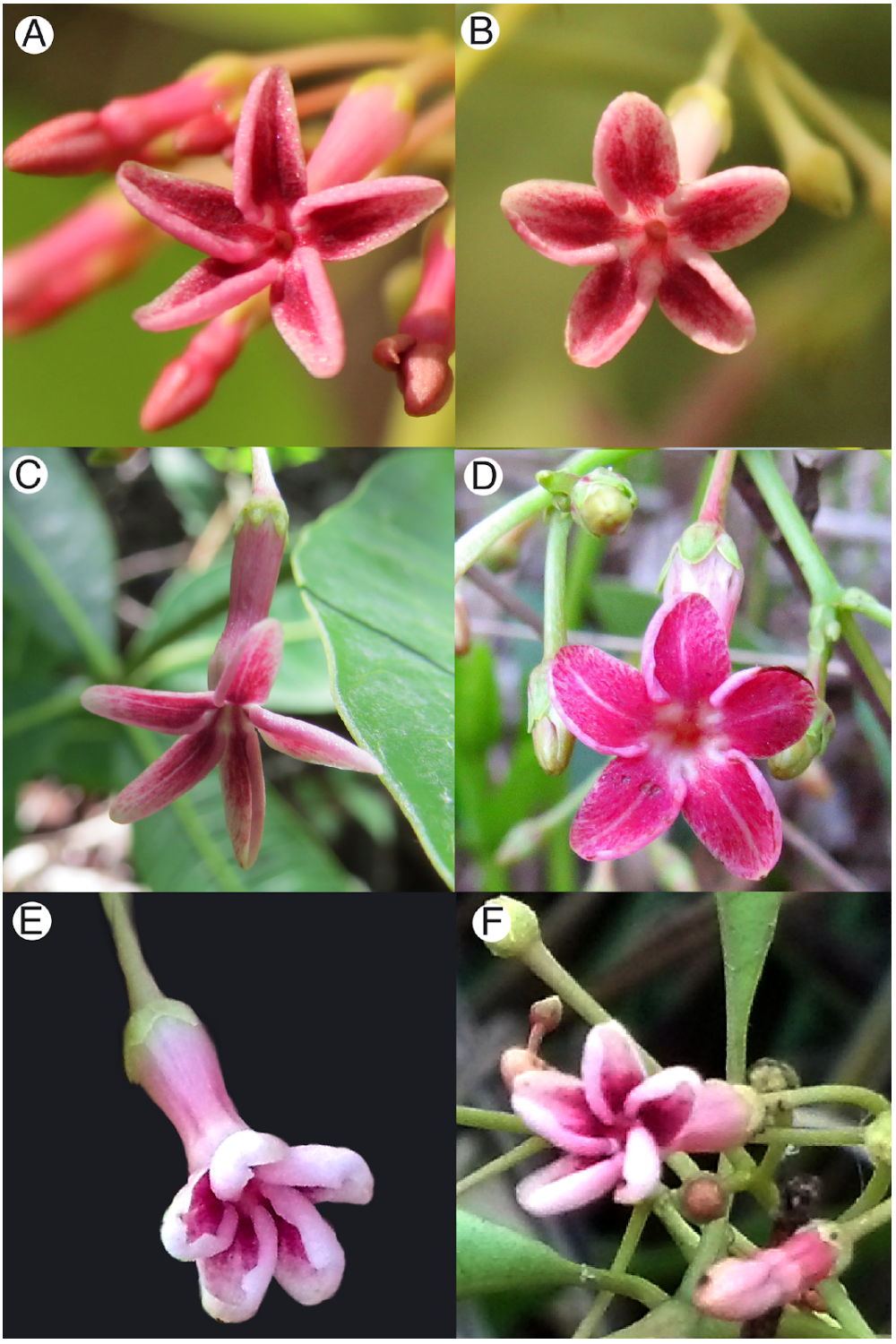


Figura 1: Flores de *Rauvolfia weddelliana* Müll.Arg. (A, B, C, D) e *Rauvolfia gracilis* I.Koch & Kin.Gouv (E, F) de diferentes localidades (A, B: Mato Grosso; C, D: Goiás; E, F: Rondônia). Créditos das imagens: Fotografias A e B: Letícia Souto; C, D, E e F: João de Deus Vidal Jr.

Neste estudo nós buscamos aplicar uma abordagem integrativa de marcadores moleculares e modelagem de distribuição potencial para testar as predições da teoria dos

refúgios sobre as populações de *Rauvolfia* no Cerrado. O primeiro capítulo desta tese está em processo de revisão na revista *Brazilian Journal of Botany*, e descreve o desenvolvimento e caracterização de um conjunto de marcadores moleculares do tipo microssatélite para aplicação em estudos populacionais para *R. weddelliana*. Neste capítulo, também testamos a transferabilidade deste marcador para *R. gracilis*. Já no segundo capítulo, a ser submetido para a revista *Journal of Biogeography*, nós desenvolvemos modelos de distribuição potencial para *R. weddelliana* e aplicamos o conjunto de marcadores microssatélites gerados no capítulo anterior, em conjunto com sequências de DNA cloroplastidial e espaçadores nucleares, para testar as predições da teoria dos refúgios nestas populações. O terceiro capítulo será submetido para a revista *Diversity and Distributions* e testa as predições de três hipóteses biogeográficas propostas na literatura nos padrões de endemismo encontrados atualmente no Cerrado. Neste capítulo, nós também reconstruímos a distribuição histórica de vegetações de chapadas e matas de vales através de modelos de nicho. Nossos resultados contribuem para a melhor compreensão de como estas espécies responderam ao clima ao longo do Quaternário e contribuirão nas discussões sobre a história biogeográfica do Cerrado. Uma melhor compreensão sobre a distribuição da diversidade biológica e dos processos que moldaram sua configuração espacial possibilitarão, por exemplo, a delimitação mais eficiente de áreas de conservação e direcionamentos de coletas para futuros trabalhos de campo.

Referências

Ab'Sáber A.N. (1981) Domínios morfoclimáticos atuais e Quaternários na região dos Cerrados. *Craton & Intracraton,* **14**, 1-37.

Ab'Sáber A.N. (1983) O domínio dos cerrados: introdução ao conhecimento. *Revista do Serviço Público*, **111**, 41–55.

Barbosa A.C. de O.F., Collevatti R.G., Chaves L.J., Guedes L.B.S., Diniz-Filho J.A.F., & Telles M.P. de C. (2015) Range-wide genetic differentiation of *Eugenia dysenterica* (Myrtaceae) populations in Brazilian Cerrado. *Biochemical Systematics and Ecology*, **59**, 288–296.

Bonatelli I.A.S., Perez M.F., Peterson A.T., Taylor N.P., Zappi D.C., Machado M.C., Koch I., Pires A.H.C., & Moraes E.M. (2014) Interglacial microrefugia and diversification of a cactus species complex: Phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. *Molecular Ecology*, **23**, 3044–3063.

Caetano S., Prado D., Pennington R.T., Beck S., Oliveira-Filho A., Spichiger R., & Naciri Y. (2008) The history of Seasonally Dry Tropical Forests in eastern South America: Inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Molecular Ecology*, **17**, 3147–3159.

Cerón-Souza I., Bermingham E., McMillan W., & Jones F. (2012) Comparative genetic structure of two mangrove species in Caribbean and Pacific estuaries of Panama. *BMC Evolutionary Biology*, **12**, 205.

Collevatti R.G., Rabelo S.G., & Vieira R.F. (2009) Phylogeography and disjunct distribution in *Lychnophora ericoides* (Asteraceae), an endangered cerrado shrub species. *Annals of botany*, **104**, 655–64.

Collevatti R.G., de Castro T.G., Lima J. de S., & Telles M.P. de C. (2012) Phylogeography of *Tibouchina papyrus* (Pohl) toledo (Melastomataceae), an endangered tree species from

rocky savannas, suggests bidirectional expansion due to climate cooling in the Pleistocene. *Ecology and Evolution*, **2**, 1024–1035.

Fiaschi P. & Pirani J.R. (2009) Review of plant biogeographic studies in Brazil. *Journal of Systematics and Evolution*, **47**, 477–496.

Haffer J. (1969) Speciation in Amazonian Forest Birds. *Science*, **165**, 131–137.

Hamada H., Petrino M.G., & Kakunaga T. (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **79**, 6465–6469.

Jarne P. & Lagoda P.J.L. (1996) Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution*, **11**, 424–429.

Koch, I. (2002). Estudos das espécies neotropicais do gênero *Rauvolfia* L. (Apocynaceae). Tese de doutorado. Universidade Estadual de Campinas, Campinas, São Paulo, Brasil.

Koch I., Kinoshita L.S., & Bittrich V. (2007) Taxonomic Novelties in *Rauvolfia* (Apocynaceae, Rauvolfioideae) from Brazil. *Novon: A Journal for Botanical Nomenclature*, **17**, 462–471.

Levinson G. & Gutman G. a (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular biology and evolution*, **4**, 203–221.

Lewis W.H. (1980) Polyploidy in Angiosperm: dicotyledons. *Polyploidy: Biological Relevance Plenum Press*, 241–268.

Peres E.A., Sobral-Souza T., Perez M.F., Bonatelli I.A.S., Silva D.P., Silva M.J., & Solferini V.N. (2015) Pleistocene niche stability and lineage diversification in the subtropical spider *Araneus omnicolor* (Araneidae). *PLoS ONE*, **10**, 1–20.

Prado C.P.A., Haddad C.F.B., & Zamudio K.R. (2012) Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology*, **21**, 921–941.

Raghavan R. (1957) Chromosome numbers in Indian medicinal plants. *Proceedings of the Indian Academy of Sciences*, **45**, 294–298.

Rao A. (1956) A revision of *Rauvolfia* with particular reference to the American species. *Annals of the Missouri Botanical Garden*, **43**, 253–354.

Rapini A., Koch I., & Simões A. (2010) Rauvolfia anomala, uma nova espécie de Apocynaceae da Chapada dos Guimarães, Mato Grosso, Brasil. *Rodriguésia*, **61**, 95–100.

Ratter J.A. & Dargie T.C.D. (1992). An analysis of the floristic composition of 26 cerrado areas in Brazil. *Edinburgh Journal of Botany*, *49*(2), 235-250.

Ratter J. a., Ribeiro J.F., & Bridgewater S. (1997) The Brazilian Cerrado Vegetation and Threats to its Biodiversity. *Annals of Botany*, **80**, 223–230.

Santos M.G., Nogueira C., Giugliano L.G., & Colli G.R. (2014) Landscape evolution and phylogeography of Micrablepharus atticolus (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography*, **41**, 1506–1519.

Turchetto-Zolet A.C., Pinheiro F., Salgueiro F., & Palma-Silva C. (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–1213.

Vanzolini P.E. & Williams E.E. (1981). The vanishing refuge: a mechanism for ecogeographic speciation. *Papéis Avulsos de Zoologia*, *34*(23), 251-255.

Vieira F. de A., Novaes R.M.L., Fajardo C.G., Santos R.M. dos, Almeida H. de S., Carvalho D. de, & Lovato M.B. (2015) Holocene southward expansion in seasonally dry tropical forests in South America: phylogeography of *Ficus bonijesulapensis* (Moraceae). *Botanical Journal of the Linnean Society*, **177**, 189–201.

Werneck F.P. (2011) The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews*, **30**, 1630–1648.

2. Development and cross-validation of microsatellite markers for *Rauvolfia weddelliana* Müll.Arg. (Apocynaceae) species complex

João de Deus Vidal Jr.[1]*, Maria Beatriz Cortez[2], Fábio Matos Alves[3], Samantha Koehler[2], Anete Pereira de Souza[3], Ingrid Koch[2]

[1]Universidade Estadual Paulista "Júlio de Mesquita Filho" – Instituto de Biociências de Botucatu -Departamento de Botânica. Rua Prof. Dr. Antônio Celso Wagner Zanin, s/nº. Rubião Júnior – Botucatu - SP. CEP 18618-689.

[2]Universidade Estadual de Campinas, Departamento de Biologia Vegetal. Rua Monteiro Lobato, 255. Cidade Universitária "Zeferino Vaz" - Barão Geraldo - Campinas – SP. Caixa Postal 6109 CEP 13083-970.

[3]Universidade Estadual de Campinas, Centro de Biologia Molecular e Engenharia Genética (CBMEG). Av. Cândido Rondon, 400. Cidade Universitária "Zeferino Vaz" - Barão Geraldo - Campinas – SP. Caixa Postal 6010 CEP: 13083-875.

*Email: jdvidal@ibb.unesp.br; Phone number: +55 15 99670-3193

ORCID id: João de Deus Vidal Jr. - 0000-0002-0420-4102; Samantha Koehler - 0000-0003-0955-5339; Ingrid Koch - 0000-0003-3256-5922.

Running title: Development of microsatellite markers for *Rauvolfia weddelliana* Müll.Arg.

Abstract

Rauvolfia weddelliana is an endemic species restricted to plateau landscapes in South American savannahs. Rapid loss of habitat and expansion of agriculture in Central-West Brazil have critically reduced the original extent of savannahs, representing a major threat to

its biotic diversity. Due to the discontinuous distribution of *R. weddelliana* and the vulnerability of its habitats, it is crucial to estimate the genetic diversity of remaining populations. The application of microsatellite markers is a useful approach with relative low cost and high informative potential for studies related to conservation genetics and population genetics. The development of specific libraries for endangered species may aid future studies about the connectivity of populations, reproductive biology and genetic diversity. We developed microsatellite markers for *R. weddelliana* and tested the transferability of the markers to a closely related species, *R. gracilis*. Ten microsatellite markers were identified and a set of primers for their amplification is presented. Most identified motifs were dimers, with lengths from 18 to 74 base pairs. Nine markers presented high informative potential (PIC > 0.5). The set of markers developed in this study will support further investigations in population genetics of *R. weddelliana* and possibly of closely related species.

Keywords: Cerrado; Conservation genetics; Population genetics; SSR.

Introduction

Rauvolfia L. is a genus of the Apocynaceae family widely distributed in tropical regions, with ca. 60 species, being 20 of them recorded for Brazil (Koch et al. 2015). Among the Brazilian species, *R. weddelliana* Müll.Arg. has the wider non-continuous distribution within the Brazilian savannahs (i.e., "Cerrados"), occurring frequently on sandy patches, often associated with plateaus in Central-West region (Rao, 1956; Koch et al. 2007). Its geographical distribution is recorded from Chapada dos Guimarães to the Paraná Basin, reaching northeastern Paraguay in isolated patches. *Rauvolfia weddelliana* forms a taxonomic complex with two other Brazilian species that are very similar in terms of ecology and/or morphology: *Rauvolfia anomala* I.Koch & Kin.-Gouv., which co-occurs in Chapada dos Guimarães and *Rauvolfia gracilis* I.Koch & Kin.-Gouv., restricted to Chapada dos Parecis, in north-western Brazil state of Rondônia (Koch et al. 2007).

Although the Cerrado constitutes the second largest domain in South America (Pennington et al. 2009), it is very vulnerable in terms of conservation, being pointed out as a conservation hotspot, mostly due to its high levels of endemism and exceptional loss of habitat (Myers et al. 2000). Originally presenting almost 2,000,000 km² of extension, it is estimated that more than 80% of Cerrado's original coverage was lost due to human activity, and less than 7% of the remaining vegetation can be considered legally protected (Myers et al. 2000; Silva & Bates, 2002). Most of *R. weddelliana* distribution is placed within Mato Grosso and Mato Grosso do Sul, core regions of extensive agronomic activity in Brazil. Although *R. weddelliana* is not listed as an endangered species (The IUCN Red List of Threatened Species. Version 2016-3. <www.iucnredlist.org>), this scenario of habitat vulnerability makes the assessment of its remaining genetic diversity a fundamental evaluation of the actual risk it

is facing, since most of its remaining area of distribution can be degraded or even disappear in a very close future. Moreover, only a few populations of *R. weddelliana* have been registered outside conservation units in the last decade, placing as urgent an estimation of how well preserved these remaining populations are in terms of genetic diversity and connectivity.

Microsatellite markers (or Simple Sequence Repeats – 'SSRs') are among the most traditional approaches in demographic studies, being used for exploring several aspects of genetic diversity with great informative potential (e.g. Zalapa et al. 2012; Hodel et al. 2016). Microsatellites are short (1-6 bp), tandemly repeated DNA sequences motifs found randomly throughout the genome of all eukaryotes (Hamada *et al*. 1982). Since SSR markers can provide important insights about population structure, gene flow and genetic diversity, they have become a major trend in studies focused on genetic characterization of natural populations, genome mapping, and parentage analysis (e.g. Hodel *et al*. 2016). Microsatellite markers are valuable tools for phylogeographic and population genetic studies, especially due to their co-dominant inheritance and relative low cost for genotyping, when compared to SNP markers, for example.

The major drawback in the usage of microsatellite markers used to be the development of libraries, which involves several steps and substantial laboratory effort (Squirrell et al. 2003). However, recent advances in sequencing technologies and improvements on methodology have allowed for a more accessible use of these markers in non-model species (Zalapa et al. 2012). Because of its high informative power and relative cost efficiency, microssatelite markers became very popular among studies in population genetics and are currently a major choice in the field (e.g. Hodel et al. 2016). Advantages involved in adopting the use of SSRs include their abundance, reproducibility, co-dominant heritage, multi-allelic

nature and high coverage of the genome (Powell et al. 1996; Kalia et al. 2011, Hodel et al. 2016). Microsatellites are usually transferable between closely related species. In fact, more than 70% of microsatellite markers designed for eudicot species amplified positively in sister species (Barbará et al. 2007). Of these, 10% were also transferrable to closely related genera (Barbará et al. 2007). Therefore, microsatellites have a potential application in multispecies studies regardless limitations due to high interspecific polymorphism.

Despite their popularity, only a few studies involving microsatellite markers have been developed for the tribe Vinceae from the Rauvolfioid grade of the Apocynaceae, mostly directed to *Catharanthus roseus* (L.) G.Don (Shokeen et al. 2007). So far, there has been no account of any similar study with other groups of the Rauvolfioid grade. In the present study, we fill in this gap. We developed and applied a set of SSR markers for *R. weddelliana* and estimated diversity related parameters, as a first step to characterize the genetic diversity and structure of populations within this species.

Material and methods

Population sampling and DNA extraction - We sampled an individual of *R. weddelliana* in Rondonópolis, Mato Grosso (12°31'31"S/60°23'09"W), collecting young leaves in silica, which were later stored under 4ºC. We deposited voucher material for the sample in the herbarium UEC (BRAZIL, Mato Grosso: Rondonópolis, *18 km a norte do município, nas margens da BR 364*, *sentido* Jaciara, 12°31'31"S, 60°23'09"W, 16-IX-2015, *Vidal, J.D.* 197639 (UEC) "[BRAZIL, Mato Grosso State: Rondonópolis (city), 18 km north from the city, along route BR 364, Jaciara (city) bound, 12°31'31"S, 60°23'09"W, 16-IX-2015, Vidal, J.D. 197639 (UEC)]"). We extracted total DNA with Qiagen DNeasy Plant Mini (Quiagen,

Valencia, CA, USA). In addition, 70 other samples from different populations along the distribution of the species (Table 1) were genotyped to characterize the polymorphism of the recovered loci. We also genotyped a population of eight individuals of *Rauvolfia gracilis* to test for primer transferability.

Microsatellite library development - We developed an enriched library of SSR following the protocol described by Billotte et al. (1999). The first step consisted in the fragmentation of extracted DNA with the restriction enzyme *Rsa*I, in order to generate fragments of lengths between 280-600 bps. We then hybridized these fragments with biotinylated oligonucleotides complementary to repetitive sequences $(CT)_8$ and $(GT)_8$, aiming to select SSR rich fragments. Once hybridized, we were able to recover these fragments with magnetic streptavidin-coated beads (Streptavidin MagneSphere Paramagnetic Particles, Promega, Madison, WI). To develop enriched libraries, we inserted recovered fragments into plasmid vectors and controlled their replication with x-gal and IPTG indicators. Positive clones (*i.e.*, non-recombining colonies) were submitted to an alkaline lysis in order to recover plasmidial DNA. Next, we sequenced the recovered inserts in an automatic sequencer ABI3500 XL Genetic Analyzer (Applied Biosystems, Foster City, CA, EUA) and analyzed sequences using the software Geneious 9.0 (http://www.geneious.com, Kearse et al. 2012). We trimmed low-quality ends and vector contamination from sequences using Geneious plugin of NCBI's VecScreen (http://www.ncbi.nlm. nih.gov/VecScreen/VecScreen.html). We screened trimmed sequences for tandem repeats with Phobos plugin (Mayer, 2010). For amplification of recovered regions, we designed complementary primers to flanking regions using the software Primer3 (Rozen & Skaletsky, 2000). To optimize genotyping, we adopted the following

parameters for primer design: 18 to 22 bp primers; T_m between 45˚C and 65˚C, with up to 3˚C of difference between each oligo in a pair of primers; salt concentration of 50 mM; GC content between 40% and 60%; and fragments' length between 100 and 360 bp. We deposited resulting sequences and annotations in GenBank (MF447169-MF447178).

Polymerase Chain Reaction (PCR) - We amplified the selected regions through Polymerase Chain Reaction (PCR) according to the following protocol: PCR mix was as follows: 50 mM KCl, 20 mM Tris–HCl (pH 8.4), 2 mM $MgCl_2$, 0.2 mM dNTP, 0.19 mg/mL BSA (bovine serum albumin), 0.15 mM of each primer, 1 U of *Taq* DNA polymerase, 2.5 µg template DNA. We completed final volume of reactions to 10 µL with ultrapure water. We applied the following temperature cycle: initial denaturing at 94˚C for 5 min, followed by 30 cycles of 94˚C (1 min), Tm 62˚C (1 min), 72˚C (2 min), and final elongation at 72˚C for 5 min.

Polymorphism evaluation - We amplified all ten recovered loci for 71 individuals of different populations of *R. weddelliana* and eight individuals of *R. gracillis* to estimate their polymorphism levels. Once amplified, we manually evaluated the polymorphism of these products through electrophoresis in polyacrylamide 6% gel stained with silver nitrate (Creste et al. 2001).

Also, since no information about ploidy was currently available for neither *R. weddelliana* nor *R. gracillis*, we analyzed band patterns for all loci to indirectly assess ploidy of species. We calculated basic allelic richness parameters (number of alleles, observed and expected heterozygosity) with adegenet R-package (Jombart, 2008). Polymorphism information content (PIC) was calculated with R package PopGenKit (Paquette, 2012),

according to the formulae proposed by Botstein et al. (1980). We also calculated F-statistics per loci (Weir & Cockerham, 1984) with R-package pegas (Paradis, 2010) for *R. weddelliana.* F-statistics were not calculated for *R. gracilis* due to limited sample (eight individuals from a single population). Package pegas was also applied to test deviations from Hardy-Weinberg Equilibrium (HWE). We calculated two estimates for HWE: chi-square test based on allele frequencies and exact test with Monte Carlo permutation of alleles (1000 repetitions). We also calculated frequency estimates of null alleles (Dempster et al. 1977) with FreeNA software (Chapuis & Estoup, 2007).

Results and discussion

We recovered a total of ten microsatellite markers from a set of 200 sequenced clones (Table 2). Motif recovery in clone amplification (5%) was higher than the average ratio in plants (2.3%; Zane *et al* 2002). Nine recovered loci were composed of dinucleotides and one was made of a tetra-/dinucleotide motif (SSR02). The occurrence of non-exclusively dinucleotide microsatellites is an expected product of enriched-libraries since probe hybridization with heterogeneous microsatellite regions is expected (Refseth *et al.* 1997). All loci were polymorphic and showed consistent amplifications following optimization. Microsatellite lengths varied from 18 to 74 base pairs, GT being the most common motif (60% of the motifs). The lengths of the amplified products ranged from 126 (SSR05) to 286 (SSR10) base pairs. Primers designed for *R. weddelliana* can be transferred between to *R. gracilis* samples, as we recovered positive cross-amplifications for all of them. All ten amplified markers were also polymorphic for both species. The number of alleles per locus

ranged from five to 27, with an average of 14 and 5 alleles per locus in *R. weddelliana* and *R. gracilis*, respectively (Table 3).

Expected heterozygosity (H_E) values for each locus in *R. weddelliana* ranged from 0.52 to 0.93 (mean value = 0.76), while observed heterozygosity (H_O) ranged from 0.34 to 0.82 (mean value = 0.55, Table 3). We also observed a deficit of heterozygotes observed for most loci ($F_{IS} > 0$, except SSR04 and SSR09). On contrast, negative F_{IS} values observed for SS04 and SSR09 represent an excess of heterozygotes in comparison with the expectation under HWE. This latter pattern may be consistent with a differential survival advantage observed for heterozygotes in plants, especially for small populations (Lesica & Allendorf, 1992). We identified deviation in HWE (Table 4) in locus SSR09, both for the chi-square (p = 0.8) and exact tests (p = 0.67).

Each individual presented up to two alleles, which strongly suggests that both species are diploid, as previously reported for other *Rauvolfia* species (Carr, 1978; Lewis, 1980; Banerjee & Sharma, 1989). Diploidy was reported for the Asian species *Rauvolfia serpentina* (L.) Benth. ex Kurz, the African species *Rauvolfia vomitoria* Afzel. (Banerjee & Sharma, 1989) and the Hawaiian species *Rauvolfia sandwicensis* A.DC. (Carr, 1978). Tetraploid and hexaploid species were also described for *Rauvolfia verticillata* (Lour.) Baill. and *Rauvolfia tetraphylla* L., respectively (Raghavan, 1957). This study provides the first estimation of ploidy for Brazilian species of the genus.

Polymorphism information content varied from 0.22 (SSR08) to 0.83 (SSR03) (Table 3). The most informative locus (SSR02; PIC=0.931) was also the one with largest repetition size, which is associated with higher mutation rate for microsatellites (Schug *et al*. 1998, Bhargava & Fuentes 2010). Higher frequencies of null alleles (29%) were observed for locus

SSR05 in Serra de São Vicente population (SV) and for loci SSR06 and SSR08 in Rondonópolis population (RP). The presence of null alleles may impact estimates of genetic diversity and population structure (Chakraborty *et al*. 1992). Due to mutations in the annealing site of primers, some alleles fail to amplify in PCR, resulting in a null allele (Chapuis & Estoup 2007). Nonetheless, it is also important to point that null alleles may also result from amplification errors, usually related to template quality (Foucault *et al*. 1996).

Most population parameters were consistent with a scenario of reduced gene flow. Populations of *R. weddelliana* and *R. gracilis* indeed present low densities of individuals and reduced sizes, suggesting demographic stochasticity may play a major role in defining the structure of genetic diversity (Schaal & Leverich 1996, Honnay & Jacquemyn 2007). However, genetic diversity parameters were based on a reduced sample size and, therefore, require careful interpretation as low sample size may obscure their biological meaning. Statistics here presented are preliminary and just an illustration of the potential of the new markers. Complementary population genetic studies are being developed to elucidate key historical demographic processes within *R. weddelliana* as well to contribute with conservation policies for the Cerrado.

Availability of data and material

The data set supporting the results of this article are included within the article supplementary data.

The testimony sample for the individual used in the library development was deposited at the Herbarium Universidade Estadual de Campinas (UEC - Campinas, SP, BR) and registered as *R. weddelliana* - 197639 (Rondonópolis, MS).

Authors' contributions

JDV conceived the study, sampled specimens in the field, performed laboratory procedures, analyzed data

and wrote the manuscript. MBC assisted with laboratory procedures and the elaboration of the manuscript. FMA assisted with laboratory procedures and gave a major contribution in data analysis and discussion of results. SK assisted with the study design, laboratory procedures and the elaboration of the manuscript. APS assisted with supervision of laboratory procedures, provided methodological support and laboratory structure for the development of the libraries. IK contributed with the collection of specimens in the field and the elaboration of the manuscript.

Acknowledgements

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) in the form of a scholarship to J.D. Vidal. Fundo de Apoio ao Ensino, à Pesquisa e Extensão – FAEPEX (process 519.292) granted funding for library sequencing and field trips to Goiás. Programa de Apoio à Pós-Graduação – PROAP, from Universidade Estadual Paulista "Júlio de Mesquita Filho" granted funding to field trips to Mato Grosso do Sul, Mato Grosso and Rondônia. This research was also supported by CAPES - Computational Biology Program.

We would like to thank professors Adarilda Benelli & Luzia Francisca de Souza for supporting our fieldwork in Mato Grosso and Goiás, respectively, and Danilo Augusto Sforça & Aline da Costa Lima Moraes for assisting daily laboratory routines in LAGM.

References

Banerjee N & Sharma AK. 1989. Chromosome constitution and alkaloid content in *Rauwolfia* L. (Apocynaceae). *Cytologia* 54: 723–728.

Barbará T, Palma-Silva C, Paggi GM, Bered F, Fay MF, & Lexer C. 2007. Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Mol Ecol* 16: 3759–3767.

Bhargava A & Fuentes FF. 2010. Mutational dynamics of microsatellites. *Mol Biotechnol*, **44**, 250–266.

Billotte N, Lagoda, PJR, Risterucci AM & Baurens FC. 1999. Microsatellite enriched libraries: applied methodology for the development of SSR markers in tropical crops. *Fruits* 54: 277–288.

Botstein D, White RL, Skolnick M & Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32: 314–331.

Carr GD. 1978. Chromosome numbers of Hawaiian flowering plants and the significance of cytology in selected taxa. *Am J Bot* 65: 236–242.

Chakraborty R, Andrade M de, Daiger SP, & Budowle B. 1992. Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Ann Hum Genet*, **56**, 45–57.

Chapuis M-P & Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* 24: 621-631.

Creste S, Neto AT & Figueira A. 2001. Detection of single sequence repeat polymorphisms in denaturing polyacrylamide sequencing gels by silver staining. *Plant Mol Biol Rep* 19: 299–306.

Dempster AP, Laird NM & Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc B Met* 39: 1–38.

Foucault F, Praz F, Jaulin C, & Amor-Gueret M. 1996. Experimental limits of PCR analysis of (CA)n repeat alterations. *Trends Genet*, **12**, 450–452.

Hamada H, Petrino MG & Kakunaga T. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *P Natl Acad Sci Usa* 79: 6465–6469.

Hodel RGJ, Segovia-Salcedo MC, Landis JB, Crowl AA, Liu X, Gitzendanner MA, Douglas NA, Germain-Aubrey CC, Chen S, Soltis DE, & Soltis PS. 2016. The Report of My Death was an Exaggeration: A Review for Researchers Using Microsatellites in the 21st Century. *Appl Plant Sci* 4: 1–13.

Honnay O & Jacquemyn H. 2007. Susceptibility of common and rare plant species to the genetic consequences of habitat fragmentation. *Conserv Biol*, **21**, 823–831.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24: 1403–1405.

Kalia RK, Rai MK, Kalia S, Singh R & Dhawan aK. 2011. Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177: 309–334.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Mentjies P & Drummond A. 2012.

Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.

Koch I, Kinoshita LS & Bittrich V. 2007. Taxonomic novelties in *Rauvolfia* (Apocynaceae, Rauvolfioideae) from Brazil. *Novon* 17: 462–471.

Koch I, Rapini A, Simões, AO, Kinoshita LS, Spina AP, Castello ACD. 2015. Apocynaceae in Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro. Available in: <http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB4839>. Accessed in: 20 jan. 2017. BFG. Growing knowledge: an overview of Seed Plant diversity in Brazil. *Rodriguésia* 66(4): 1085-1113.

Lesica P & Allendorf FW. 1992. Are small populations of plants worth preserving? *Conserv Biol*, **6**, 135–139.

Lewis WH. 1980. Polyploidy in Angiosperm: dicotyledons. *Polyploidy: Biological Relevance* Plenum Press: 241–268.

Mayer C. 2010. Phobos - a tandem repeat search tool for complete genomes. Version 3.3.12. Available in: <http://www.rub.de/spezzoo/cm/cm_phobos.htm> Accessed in: 20 jan. 2017.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, & Kent J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853–858.

Paquette SR. 2012. PopGenKit: Useful functions for (batch) file conversion and data resampling in microsatellite datasets. *R package version*, *1*.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* 26: 419-420.

Pennington RT, Lavin M & Oliveira-Filho A. 2009. Woody plant diversity, evolution, and ecology in the Tropics: perspectives from seasonally dry tropical forests. *Annu Rev Ecol Evol S* 40: 437–457.

Powell W, Machray GC & Provan J. 1996. Polymorphism revealed by simple sequence repeats. *Trends Plant Sci* 1: 215–222.

Raghavan R. 1957. Chromosome numbers in Indian medicinal plants. *P Indian Acad Sci B* 45: 294–298.

Rao A. 1956. A revision of *Rauvolfia* with particular reference to the American species. *Ann Mo Bot Garden* 43: 253–354.

Refseth UH, Fangan BM, & Jakobsen KS. 1997. Hybridization capture of microsatellites directly from genomic DNA. *Electrophoresis*, **18**, 1519–1523.

Rozen S. & Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S. & Misener S. (eds) Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa, NJ: 365-386.

Schaal BA & Leverich WJ. 1996. Molecular variation in isolated plant populations. *Plant Spec Biol*, **11**, 33–40.

Schug MD, Hutter CM, Wetterstrand KA, Gaudette MS, Mackay TF, & Aquadro CF. 1998. The mutation rates of di-, tri- and tetranucleotide repeats in *Drosophila melanogaster*. *Mol Biol Evol*, **15**, 1751–1760.

Shokeen B, Sethy NK, Kumar S & Bhatia S. 2007. Isolation and characterization of microsatellite markers for analysis of molecular variation in the medicinal plant Madagascar periwinkle (*Catharanthus roseus* (L.) G. Don). *Plant Sci* 172: 441–451.

Silva JMC & Bates JM. 2002. Biogeographic patterns and conservation in the South American Cerrado: A tropical savanna hotspot. *BioScience* 52: 225.

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, & Powell W. 2003. How much effort is required to isolate nuclear microsatellites from plants? *Mol Ecol* 12: 1339–1348.

Weir BS & Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358.

Zalapa J.E., Cuevas H., Zhu H., Steffan S., Senalik D., Zeldin E., McCown B., Harbut R. & Simon P. 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am J Bot* 99: 193–208.

Zane L, Bargelloni L, & Patarnello T. 2002. Strategies for microsatellite isolation: a review. *Mol Ecol*, **11**, 1–16.

Tables and figures

Table 1: Sampled populations of *Rauvolfia weddelliana* Müll.Arg. and respective localities. *Locality*: Sampling locality. *Population*: Name attributed to the population; *ID*: Abbreviation of population names. *N*: number of individuals sampled on population. (*): Location placed inside conservation units. (**) *Rauvolfia gracilis* I.Koch & Kin.-Gouv. population.

Locality	Population	ID	N	Latitude	Longitude
Mato Grosso, Chapada dos Guimarães*	Faz. Chafariz	CP	14	15°17'54"S	55°50'37"W
Mato Grosso, Chapada dos Guimarães*	Véu de Noiva	VN	29	15°24'36"S	55°49'50"W
Mato Grosso, Serra de São Vicente	Jaciara	SV	5	15°49'06"S	55°20'38"W
Mato Grosso, Rondonópolis	Rondonópolis	RP	5	12°31'31"S	60°23'09"W
Goiás, PARNA das Emas*	Mineiros	CD	10	18°17'35"S	52°53'44"W
Goiás, RPPN Pousada das Araras*	Serranópolis	PA	10	18°27'03"S	52°00'22"W
Rondônia, Vilhena**	Vilhena	VL	8	18°27'03"S	52°00'22"W

Table 2: Set of microsatellite markers recovered for *Rauvolfia weddelliana* Müll.Arg. with respective primers and properties. *Locus*: name attributed to the locus; *Sequence*: sequence of bases of the primer (5'->3' sense - *F*: forward; *R*: reverse); *%GC*: percentage of G (guanine) and C (cytosine) in the primer composition. T_M: primer melting temperature; *H*: primer hairpin; *P.size*: size of the whole amplified fragment; *Motif*: microsatellite basic unity of repetition and number of repeats.

Locus	Sequence (5'->3')	% GC	T_M (ºC)	H (ºC)	P.size (bp)	Motif	GenBank accession
SSR01	F: ACAGGAGTGTCAAAATCCAA	40	54.8	-	250	$(GT)_9$	MF447171
	R: CTTGTTTCGAGGCAGTGATG	50	57.1	-			
SSR02	F: AGAAATGCGTATCCAATGCG	45	57	-	210	$(GTGA)_4(GA)_{19}(CA)_{10}$	MF447170
	R: AAAGATGTCAGGTCCCACTG	50	57.1	-			
SSR03	F: ATCTATTCTCCAGCCTGTGC	50	57.1	38.0	220	$(TG)_9$	MF447173
	R: GGCCCTAACAATTGGTCTCT	50	57	-			
SSR04	F: GCACACCATACACTGCTCTA	50	57.3	-	151	$(TG)_8(GT)_9$	MF447174
	R: GAGGTCAAAAAGCTGTTCCC	50	56.9	35.3			
SSR05	F: TTTCCAAAGCTGCCTCAAAG	45	56.8	-	126	$(AC)_7$	MF447169
	R: AAACATGGTTCTCACACCCT	45	56.9	35.7			
SSR06	F: GAGTGTGGAACCTGTCATGA	50	57.2	-	252	$(AC)_{15}$	MF447177
	R: GCTCCTGATGTCTGTTCAGA	50	57	42.6			
SSR07	F: AACAGCCCCTTCATCATCAA	45	57.1	-	210	$(GT)_8$	MF447175
	R: GGACAAGTTTTTCTCCCTGC	50	56.9	-			
SSR08	F: AGGCTGAAAGTAACGACTGA	45	56.5	35.5	144	$(TG)_8$	MF447172

	R: TCTGTCTCTCAGTCCCAGAA	50	57	44.0			
SSR09	F: ACCTCCGTAATTGTGGAACA	45	56.8	32.9	161	$(AC)_{13}$	MF447176
	R: CGGTTCAGGAGAGAGAAACA	50	56.9	31.0			
SSR10	F: AAAGTGCAGACTACCGACAA	45	57.1	-	286	$(GA)_{20}$	MF447178
	R: TTGAACAGTTTTTGCCGGTT	40	57	-			

Table 3: Allelic richness statistics for the microsatellite markers described for *Rauvolfia weddelliana* Müll.Arg and *R. gracilis* I.Koch & Kin.-Gouv.. *Locus*: identification of the locus; *A*: number of alleles observed per locus in eight samples used for characterization; H_E: expected heterozygosity; H_O: observed heterozygosity; F_{IS}: inbreeding coefficient; *PIC*: polymorphism information content.

	Total (n=79)				*Rauvolfia weddelliana* (n=71)				*Rauvolfia gracilis* (n=8)		
Locus	*A*	H_E	H_O	PIC	*A*	H_E	H_O	F_{IS}	*A*	H_E	H_O
SSR01	27	0.90	0.52	0.893	26	0.89	0.48	0.41	6	0.80	0.75
SSR02	27	0.94	0.74	0.931	26	0.93	0.75	0.18	10	0.88	0.75
SSR03	18	0.91	0.75	0.907	19	0.91	0.73	0.11	3	0.55	1.00
SSR04	12	0.90	0.85	0.886	12	0.89	0.82	-0.03	4	0.73	1.00
SSR05	9	0.58	0.35	0.562	9	0.54	0.34	0.35	5	0.72	0.50
SSR06	14	0.77	0.42	0.741	14	0.80	0.45	0.25	3	0.23	0.25
SSR07	6	0.74	0.44	0.696	6	0.73	0.40	0.33	3	0.60	0.71
SSR08	5	0.49	0.42	0.462	6	0.52	0.40	0.30	2	0.38	0.50
SSR09	8	0.60	0.71	0.572	8	0.55	0.65	-0.18	5	0.73	1.00
SSR10	10	0.81	0.43	0.788	10	0.81	0.44	0.45	8	0.79	0.50
Mean	14	0.76	0.56	0.744	14	0.76	0.55	-	5	0.64	0.70

Table 4: Null allele frequencies estimated for microsatellites described for *Rauvolfia weddelliana* Müll.Arg. *Locus*: identification of the locus; *Pr(chi²)*: chi-square test calculation for Hardy-Weinberg Equilibrium; *Pr (exact)*: exact test for Hardy-Weinberg Equilibrium based on 1000 Monte Carlo permutations of alleles. *SV*: population from Serra de São Vicente, Jaciara (MT); *VN*: sample from Véu de Noiva, Chapada dos Guimarães (MT); *CP*: sample from Cidade de Pedra (MT); *VL*: population from Vilhena (RO); *RP*: population from Rondonópolis (MT); *CD*: population from PARNA Emas, Mineiros (GO); *PA*: population from RPPN Pousada das Araras, Serranópolis (GO).* Locus with significant deviation from Hardy-Weinberg Equilibrium (chi-square p = 0.8; exact tests p = 0.67).

	Null allele frequency estimated per population						
Locus	SV	VL	RP	CD	PA	VN	CP
SSR01	0.24	0.04	0.00	0.00	0.00	0.24	0.23
SSR02	0.11	0.07	0.00	0.04	0.00	0.10	0.01
SSR03	0.19	0.00	0.07	0.00	0.07	0.00	0.10

SSR04	0.00	0.00	0.00	0.00	0.00	0.04	0.09
SSR05	**0.29**	0.12	0.00	0.00	0.00	0.08	0.14
SSR06	0.00	0.00	**0.29**	0.00	0.00	0.07	0.06
SSR07	0.00	0.00	0.00	0.03	0.00	0.21	0.12
SSR08	0.00	0.00	**0.29**	0.00	0.00	0.00	0.17
SSR09*	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SSR10	0.27	0.16	0.20	0.01	0.16	0.19	0.10

Supplementary data

Table 1: Genotypes observed for ten loci of SSR markers developed for *Rauvolfia weddelliana* and *R. gracilis*. *Locus*: name of the locus; *Sample Id*: identification of sample (*SV*: sample from Serra de São Vicente, Jaciara (MT); *VN*: sample from Véu de Noiva, Chapada dos Guimarães (MT); *CP*: sample from Cidade de Pedra (MT); *VL*: sample from Vilhena (RO); *RP*: sample from Rondonópolis (MT); *CD*: sample from PARNA Emas, Mineiros (GO); *PA*: sample from RPPN Pousada das Araras, Serranópolis (GO)). (*) Sample from *R. gracilis.* (NA) Samples that failed to amplify.

	Locus									
Sample Id	SSR01	SSR02	SSR03	SSR04	SSR05	SSR06	SSR07	SSR08	SSR09	SSR10
Rw_1SV	304/304	204/192	242/242	178/168	130/130	244/244	208/204	142/142	160/162	272/272
Rw_2SV	308/308	188/188	238/238	170/170	134/134	244/232	208/208	142/142	160/166	272/272
Rw_3SV	302/310	192/190	238/242	162/166	138/138	244/244	208/208	148/142	160/160	272/272
Rw_4SV	NA	NA	220/220	178/166	134/134	244/244	208/208	148/142	160/162	272/272
Rw_5SV	304/304	192/192	238/250	170/166	136/134	246/242	208/208	142/142	160/160	276/276
Rg_19VL*	296/286	216/220	222/226	160/164	130/122	234/236	208/204	142/142	162/168	276/284
Rg_20VL*	296/286	180/220	222/226	160/164	132/128	234/234	208/204	142/142	162/168	280/284
Rg_21VL*	290/270	190/212	222/226	162/168	130/128	234/234	210/210	138/142	160/162	282/282
Rg_22VL*	266/286	168/206	222/226	160/164	130/126	234/234	NA	142/142	160/162	284/284
Rg_23VL*	290/270	168/208	222/226	160/164	132/132	234/234	210/208	138/142	158/160	292/282
Rg_24VL*	290/270	192/192	222/226	160/164	128/128	232/234	210/208	138/142	158/160	284/284
Rg_25VL*	270/270	192/168	222/226	162/168	128/128	234/234	210/208	138/142	160/162	272/262
Rw_27RP	250/250	186/200	240/240	162/168	124/128	254/254	210/210	148/142	166/162	282/282
Rw_28RP	250/250	172/172	242/242	160/170	128/128	256/256	210/210	142/142	160/162	282/282
Rw_29RP	268/250	208/210	242/220	162/168	128/132	252/240	210/210	142/142	160/164	NA
Rw_30RP	NA	210/178	242/220	170/178	126/130	252/252	210/210	138/138	160/158	286/286
Rw_31RP	250/250	210/178	242/220	152/156	126/130	252/252	210/210	140/140	160/164	302/286
Rw_32CD	NA	168/162	220/220	NA	124/130	234/234	208/208	138/144	160/164	302/286
Rw_33CD	268/250	186/200	218/220	152/160	124/130	234/234	210/208	140/140	160/162	292/284
Rw_34CD	280/270	210/178	220/220	152/156	130/130	234/234	210/208	142/142	160/156	302/284
Rw_35CD	270/260	178/178	220/220	152/156	130/130	234/234	210/208	142/140	160/162	282/282
Rw_36CD	274/264	168/172	220/220	152/156	130/130	234/234	210/208	142/140	160/162	284/284

Rw_37CD	260/260	200/162	220/220	156/156	132/124	234/234	210/208	140/140	160/154	292/282
Rw_38CD	270/260	170/170	218/220	152/156	124/130	234/234	208/208	140/138	160/162	284/284
Rw_39CD	250/238	168/166	218/220	152/156	132/124	234/234	208/208	144/138	160/156	292/282
Rw_40CD	258/250	166/190	218/220	152/156	130/130	234/234	208/202	140/138	NA	292/282
Rw_41CD	300/290	168/178	228/234	152/156	130/130	234/234	202/202	142/140	NA	272/262
Rw_42PA	292/286	178/178	228/234	152/156	130/130	234/234	204/202	142/140	160/160	272/272
Rw_43PA	290/284	174/194	228/228	152/156	130/130	234/234	202/200	142/142	160/160	262/282
Rw_44PA	298/282	180/192	230/234	152/156	130/130	234/232	204/202	NA	160/160	272/272
Rw_45PA	292/290	180/164	230/226	152/156	130/130	234/234	202/202	142/142	160/162	NA
Rw_46PA	292/284	178/192	232/232	152/156	130/130	234/234	208/200	140/142	160/162	262/282
Rw_47PA	294/290	164/198	230/232	152/156	130/130	234/234	202/200	142/142	162/164	276/276
Rw_48PA	300/280	178/188	230/230	152/156	124/130	234/234	208/200	142/142	160/162	NA
Rw_49PA	290/288	166/190	230/232	152/156	124/130	234/224	208/200	140/142	160/162	292/282
Rw_6VN	300/300	188/188	242/220	NA	NA	244/244	208/208	142/142	160/154	276/276
Rw_7VN	NA	190/188	242/220	178/178	130/130	244/244	208/208	144/142	160/154	276/276
Rw_8VN	300/300	188/188	242/220	178/178	NA	244/244	208/208	142/142	160/154	276/258
Rw_9VN	296/296	208/204	244/238	166/170	130/130	244/232	208/208	NA	160/154	NA
Rw_10VN	296/296	NA	242/220	180/170	NA	242/244	208/208	142/142	160/154	NA
Rw_11VN	296/296	186/200	236/240	180/170	NA	242/232	208/208	148/142	160/154	282/276
Rw_12VN	296/280	NA	232/240	180/172	NA	244/232	208/208	142/142	160/154	282/282
Rw_50VN	290/290	188/188	230/232	152/156	130/130	NA	NA	142/142	160/162	292/282
Rw_51VN	278/278	186/186	NA	162/168	130/130	234/244	202/202	NA	NA	272/262
Rw_52VN	280/272	188/188	NA	NA	130/130	NA	NA	NA	156/160	284/284
Rw_53VN	282/278	NA	NA	162/168	130/130	234/244	204/204	NA	164/168	272/262
Rw_54VN	280/280	NA	234/250	162/168	130/130	234/244	204/200	NA	160/160	272/282
Rw_55VN	280/280	NA	234/250	162/168	130/130	234/244	204/200	NA	160/160	262/282
Rw_56VN	282/280	222/232	250/232	162/168	130/130	234/244	204/204	NA	160/160	292/282
Rw_57VN	280/280	222/198	250/256	164/170	130/130	246/254	210/204	142/140	160/160	272/282
Rw_58VN	280/280	186/198	230/242	162/168	130/130	246/246	NA	142/142	160/162	284/284
Rw_59VN	NA	NA	NA	NA	130/130	234/234	208/200	NA	160/160	282/282

Rw_60VN	290/280	186/198	248/248	164/170	NA	234/252	NA	NA	160/160	282/282
Rw_61VN	290/280	180/180	256/250	166/166	130/132	246/246	204/204	142/142	160/160	282/282
Rw_62VN	290/276	198/222	248/252	162/166	130/130	244/248	204/204	148/142	160/160	282/282
Rw_63VN	282/282	180/178	252/230	162/172	130/124	248/248	204/204	148/142	160/160	NA
Rw_64VN	280/280	198/222	250/244	162/172	130/130	244/250	204/202	148/142	158/160	NA
Rw_65VN	280/280	180/180	250/244	168/168	130/136	244/256	204/204	142/142	168/160	302/284
Rw_66VN	280/280	180/178	250/244	166/172	130/130	244/244	204/204	NA	160/160	282/282
Rw_67VN	280/280	180/180	250/244	NA	130/128	244/244	204/202	142/142	160/160	284/284
Rw_68VN	280/280	180/178	252/244	166/172	130/122	244/254	204/202	148/142	158/160	284/284
Rw_69VN	280/280	180/168	240/230	162/170	128/128	232/246	204/204	148/142	158/160	284/284
Rw_70VN	280/280	NA	238/252	162/170	130/130	NA	204/204	148/142	158/160	284/284
Rw_71VN	284/280	180/212	250/262	160/166	130/130	244/254	204/204	142/142	160/168	282/282
Rw_72CP	282/282	222/198	232/232	166/170	130/132	NA	204/194	NA	160/160	284/284
Rw_73CP	282/282	184/184	250/256	166/170	130/130	242/242	208/204	142/142	160/168	284/284
Rw_74CP	284/284	180/212	250/232	166/174	130/130	238/242	204/204	NA	160/160	284/284
Rw_75CP	284/280	180/238	250/256	168/160	130/130	248/232	204/204	142/142	160/160	302/284
Rw_76CP	284/280	184/198	250/256	166/170	130/130	242/248	204/204	NA	160/160	302/284
Rw_77VN	284/284	180/212	250/232	170/170	130/130	244/244	208/204	142/142	160/160	NA
Rw_78VN	284/284	180/248	242/260	168/160	130/122	242/254	204/204	146/144	160/160	284/292
Rw_79VN	NA	180/248	242/250	168/160	130/122	242/254	204/204	144/144	160/160	NA
Rw_13CP	NA	188/198	220/220	172/172	NA	244/232	210/208	142/142	160/160	282/282
Rw_14CP	296/286	186/186	236/220	170/170	130/130	242/242	210/208	142/142	160/158	282/282
Rw_15CP	NA	188/198	240/236	172/172	132/132	244/244	210/208	142/142	160/158	282/282
Rw_16CP	290/290	186/216	230/230	172/172	128/128	244/244	208/208	142/142	160/158	292/284
Rw_17CP	296/286	186/166	226/226	172/172	132/122	242/232	208/208	142/142	160/164	286/282
Rw_18CP	290/290	186/200	222/226	172/166	128/124	234/236	210/210	142/142	160/164	280/276

3. Phylogeography and paleodistribution of *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I.Koch & Kin.-Gouv. (Apocynaceae)

Vidal, J. D.[1]*; Perez, M. F. [2]; Souza, A. P. [3]; Koch, I. [4] and Koehler, S.[4]

[1]Universidade Estadual Paulista "Júlio de Mesquita Filho" – Instituto de Biociências de Botucatu. [2]Universidade Federal de São Carlos, Departamento de Biologia, Centro de Ciências Humanas e Biológicas [3]Universidade Estadual de Campinas, Centro de Biologia Molecular e Engenharia Genética (CBMEG). [4]Universidade Estadual de Campinas, Departamento de Biologia Vegetal.

*E-mail: jdvidal@ibb.unesp.br

Abstract

Aim Here, we investigated population genetics, phylogeography, and past distribution models of two sister species of western Brazilian Cerrado, both occurring in the arid plateaus of South America. We tested possible demographic scenarios to address the impacts of Pleistocene climate and older processes into this species complex.

Location Savannah fields of Western plateaus of Brazilian Cerrado.

Methods We analysed chloroplast spacers and nuclear microsatellite markers to describe the genetic diversity and phylogeographic structure of these populations. We also applied ecological niche modeling to reconstruct the range shifts of these species during Quaternary. Finally, we applied Approximate Bayesian Computation to test demographic scenarios of glacial retraction/expansion, glacial/interglacial refugia, northward/southward colonization, and core/peripheral colonization.

Results We recovered five nuclear clusters and eight chloroplast haplotypes. Our results indicate northwest-southeast structure of populations, both for chloroplast and nuclear markers. We found a single chloroplast haplotype exclusive for *Rauvolfia gracilis*, but most chloroplast and nuclear markers are common to *R. gracilis* and *R. weddelliana* populations from southeast of Mato Grosso. Our coalescent simulations recovered a scenario of glacial retraction and interglacial expansion for populations, which corroborates an interglacial refuge scenario.

Main Conclusions

We concluded that current distribution of lineages of *Rauvolfia weddelliana* and *Rauvolfia gracilis* are interglacial refugia. Although our niche models presented contrasting results, suggesting a retraction in potential distribution of these species during last glaciation, they also recovered the same north-south division for the interglacial scenario. We found no support for the taxonomic division between these species, and identified a cryptic lineage on the southern populations. Our results reinforce the combined role of interglacial refugia and earlier geomorphological events in the biogeographic history of Cerrado.

Keywords: *Rauvolfia gracilis*, savannah, genetic structure, gene flow, Cerrado.

Introduction

Climate played a major role in the distribution of genetic diversity within species (Haffer 1969; Hewitt 2000). Due to the variation in climatic suitability along glacial cycles, populations of plants and animals were differentially isolated in climatically stable and

unstable areas (Hewitt 1996). As climate changed, organisms responded by persisting, suffering local extinction or even expanding to newly suitable areas (Davis & Shaw 2001; Aitken *et al*. 2008). Especially during the Quaternary (2.4 mya) these biological processes significantly impacted the evolution of several lineages (van der Hammen 1974; Hewitt 2000, 2004; Jansson 2003).

Glaciations promoted southward retraction towards lower latitudes in Europe and North America and further northwards re-colonization (Hewitt 2000). Even though the effects of climatic oscillations on lineage diversification were more extensively studied in the Northern Hemisphere (Markgraf *et al*. 1995; Hewitt 2004; Médail & Diadema 2009), it is known that cycles of dryer and colder glacial climates affected other areas in a global scale. In the Southern Hemisphere, South America and Australia climates also changed considerably, with increase in aridity and decrease in temperature (e.g. Pennington *et al*. 2000). Since these regions had not experienced significant ice sheets advances (Hewitt 1996), the impact on the biota was less drastic and involved changes in elevation ranges and cycles of expansion-retraction of distributions (Markgraf *et al*. 1995; Davis & Shaw 2001). In South America, the prevalence of dry climates was not homogeneous, as areas were deeply influenced by topographic fragmentation in central South America and by the presence of glacial enclaves in the Andes (Ab'Saber 1977). As aridity increased, it supposedly promoted retraction of tropical forests, whereas open vegetation habitats presumably expanded (van der Hammen 1974; Hewitt 2004).

Despite the lack of phylogeographic studies concerning open vegetation habitats (Turchetto-Zolet *et al*. 2013), the expansion of seasonally dry forests and savannahs during dry periods is partially corroborated by phylogeographic patterns of some plants and animals

(Caetano *et al*. 2008, Collevatti *et al*. 2009, Prado *et al*. 2012) and pollen records (van der Hammen, 1974; Ledru 2002). However, several studies phylogeographic patterns defy this expansion-retraction model (Turchetto-Zolet *et al*. 2013). Studies based on phylogeographic patterns, both for animals (Santos *et al*. 2014) and plants (Vieira *et al*. 2011; Barbosa *et al*. 2015), and historical vegetation models (Werneck *et al*. 2012; Werneck *et al*. 2011) from species from the Cerrado and other open biomes (i.e. Caatinga, Chaco) suggest a scenario of vegetation retraction and fragmentation during glacial periods rather than expansion. The lack of clear patterns of response to historical climate oscillations suggests that there is need of many more studies until we will be able to clarify the response of the South American biota of open domains to glacial cycles (Turchetto-Zolet *et al*. 2013).

Areas of contraction of the distributional range of a species under colder and warmer periods are defined as glacial and interglacial refugia, respectively. The effect of glacial and interglacial refugia in a species distributional range can be inferred through the description of genetic diversity patterns based on neutral molecular markers (Stewart *et al.* 2010). The contrasting differences of population genetic diversity and phylogeographic structure observed for interglacial refugia are consequences of the difference in the relative timespan of ice ages and interglacials and to how organisms respond to oscillations in climate (Stewart *et al* 2010; Arenas *et al*. 2012).

Major glaciations dominated the global climate for very long periods of time along the last 700 kyr, while warmer periods took place each 100 kyr (Hewitt 1996). Interglacials were relatively shorter, with durations of 10 to 15 kyr, while glaciations lasted up to 100 kyr (Hewitt 1996). Therefore, species of glacial refugia went through longer and slower contraction periods than interglacial species (Stewart *et al* 2010). Since range contractions

that originated glacial refugia occurred for longer periods, the genetic diversity of populations within glacial refugia is expected to be reduced compared to populations retained in interglacial refugia (Arenas *et al* 2012). Also, considering the major evolutionary process detected by neutral molecular markers is drift, the genetic diversity among populations of different refugia is expected to be greater in glacial refugia than in interglacial ones (Arenas *et al.* 2012).

Cerrado is a wide South American domain, composed by fire-adapted xeromorphic vegetation, with strong presence of grasses (Ratter *et al*. 1997). It is an ancient formation, originated before the final Gondwana breakup and with an outstanding diversity of recently originated biota (Ratter *et al*. 1997). Ancient plateaus dominate the landscape of Cerrado, harbouring savannah-like vegetation, which includes open fields (i.e. "campos limpos"; grasslands with scattered shrubs) to cerrado *sensu stricto* (i.e. woodlands with closed scrub) (Eiten 1972; Oliveira-Filho & Ratter 2002). Inter-planaltic valleys, on the other hand, generally have heterogeneous vegetation types, combining elements of semi-deciduous, deciduous and gallery forests (Pennington *et al*. 2000; Werneck 2011). Phylogeographic patterns of widespread species like *Ficus bonijesulapensis* R.M.Castro (Vieira *et al*. 2011) *Hymenaea stigonocarpa* Hayne (Ramos *et al*. 2007), *Caryocar brasiliense* A.St.-Hil. (Collevatti *et al*. 2003), *Annona coriacea*, Mart. and *A. crassiflora* Mart. (Ribeiro *et al*. 2016) supported scenarios of glacial retraction for these taxa. An interglacial refugia scenario, on the other hand, was inferred for species occurring in high elevations, like *Tibouchina papyrus* (Pohl) Toledo (Collevatti *et al*. 2012) and *Pilosocereus aurisetus* (Werderm.) Byles & G.D. Rowley (Bonatelli *et al.* 2014). Therefore, the variability of environments and altitudinal

compartmentalization may be associated with the individual responses of each species, making it difficult to assemble a general pattern.

Despite the lack of concordance, the phylogeography of species from central and northeastern regions of Cerrado is reasonably well represented in literature (Ramos *et al*. 2007; Novaes *et al*. 2010; Collevatti *et al*. 2009, 2012; Lima *et al*. 2017). Conversely, the western region of the Cerrado is still poorly sampled and investigated (Françoso *et al*. 2016). The western Cerrado presents the higher species richness of birds and mammals within the domain and was already proposed as a priority region for conservation, due to the degree of irreplaceability of its fauna (Diniz-Filho *et al*. 2009). This region comprises three important geomorphological formations: Serra dos Caiapós, Chapada dos Parecis and Chapada dos Guimarães, the latter proposed as an area of retraction and expansion of the Cerrado during glaciations (Ab'Saber 1983).

The *Rauvolfia weddeliana* complex (Apocynaceae, Rauvolfioideae) constitutes an ideal model to evaluate the effects of glacial and interglacial climates on evolutionary events in the plateaus of the Cerrado. It comprises two savannah species, *R. weddelliana* Müll.Arg. and *R. gracillis* I.Koch & Kin.-Gouv, with similar morphology and geographical distribution (Koch *et al*. 2007). The distribution of the populations is disjoint along the central and western Brazilian plateaus, including Chapada dos Parecis, Chapada dos Guimarães, and Serra dos Caiapós. Discontinuous distributions in altitudinal environments are often associated with interglacial refugia (Bonatelli *et al*. 2014). For Cerrado, some studies focused on impacts of glaciations were held for rocky outcrops highlands, named "campos rupestres" (Moraes *et al*. 2009; Collevatti *et al*. 2012, Bonatelli *et al*. 2014), but a few studies focusing species related to plateau environments (Prado *et al*. 2012; Santos *et al*. 2014) are available.

Thus, it is still uncertain if these environments follow similar patterns of glacial expansion reported for other highlands within Cerrado.

In this study we contribute to knowledge on the Quaternary biogeographic history of the western Cerrado by describing the effect of climatic oscillations on the geographic distribution and lineage diversification of the *Rauvolfia weddeliana* complex. We estimated genetic diversity and phylogeographic structure based on microsatellites and cpDNA sequences and used ecological niche modelling to reconstruct the range shifts within this group since the Last Inter-Glacial (LIG; 120,000 ybp). Lastly, we applied multi-model inference to test the suitability of glacial/interglacial refugia and northward/southward expansion scenarios, which can be combined with niche models to improve its phylogeographic interpretation, as recommended by Richards *et al*. (2007). We predict that in a scenario of glacial refugia, divergence times should indicate a Pleistocene origin, with signals of retraction during last Glacial Maximum (LGM, ca. 21,000 ybp) and high structure of genetic diversity due to long-term isolation. Contrarily, in an interglacial refugia scenario, populations should have a more ancient origin, with longer branches and older divergence times, due to their expansion during LGM, and low genetic structure. Our demographic simulations and coalescence times of lineages may also support these scenarios.

Methods

Sampling strategy and DNA extraction

We collected leaves for 77 individuals of *Rauvolfia weddelliana* in six populations of western Brazilian Cerrado and a single population of seven individuals of *Rauvolfia gracilis* (Table 1). Species were identified based on Koch *et al*. (2007). We were unable to sample *R*.

weddelliana individuals in the southern portion of its distribution due to the extremely sparse and rare occurrence of plants. We sampled individuals at least 10 m distant from each other in order to avoid the sampling of clones or the same individual, since presence of xylopod is common among *Rauvolfia* species (Simões *et al*. 2016). We defined populations as a continuous set of individuals up to 1 Km from each other. We stored leaves in silica gel and then kept under refrigeration until DNA extraction. We extracted total DNA with the CTAB protocol (Doyle & Doyle 1987).

Table 2: Population information and basic genetic diversity data for *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I. Koch & Kin.-Gouv.. *Locality*: Municipality and State of sampled populations; *Code*: Code attributed to the population; *Nmic/Ncp*: Number of individuals genotyped for microsatellite markers and chloroplast markers; A_R: Allelic richness (for microsatellite markers). H_E/H_O: Expected and observed heterozygosity (for microsatellite markers). F_{IS}: Inbreeding coefficient (for microsatellite markers). cpDNA: Haplotype attributed to intergenic cpDNA (t*rnH*-GUG/*psbA* and *trnR*-UCU/*atpA*) concatenated sequences. (¹): Location placed inside conservation units. (²): Population presenting deviation from Hardy-Weinberg Equilibrium (*p* value = 0.021).

Locality	Cod.	Lat.	Lon.	Elev.	N_{mic}/N_{cp}	A_R	H_E/H_O	F_{IS}	*cpDNA*
MT – Cuiabá, Chapada dos Guimarães (Cidade de Pedra)¹	CP	15°18'S	55°50'W	665 m	13/12	3.86	0.68/0.5	0.27	I, IV
MT – Cuiabá, Chapada dos Guimarães (Véu de Noiva)¹	VN	15°24'S	55°50'W	650 m	27/16	3.84	0.66/0.54	0.18	I, II, IV, VI
MT - Jaciara, Serra de São Vicente²	SV	15°49'S	55°20'W	690 m	5/4	2.73	0.52/0.38	0.28	V
MT - Rondonópolis	RP	12°31'S	60°23'W	529 m	5/4	3.21	0.59/0.51	0.13	V
GO - Chapadão do Céu, PARNA das Emas¹	CD	18°17'S	52°54'W	825 m	19/14	3.75	0.6/0.55	0.07	VII, VIII
GO – Serranópolis, RPPN Pousada das Araras¹	PA	18°27'S	52°00'W	628 m	8/5	3.33	0.57/0.64	-0.12	VII
RO - Vilhena	VL	18°27'S	52°00'W	628 m	7/8	3.58	0.64/0.7	-0.09	III, V

Markers selection, amplification and genotyping

To include both nuclear and chloroplast histories in our analysis, we adopted two sets of markers: nuclear microsatellite loci (nuclear short sequence repeats, nSSR) and chloroplast intergenic spacers sequences (cpDNA). We applied ten microsatellite loci specific for *R. weddelliana* to amplify a total of 77 individuals (Table 2). We manually genotyped nSSR through electrophoresis in 6% polyacrylamide gel stained with silver nitrate ($AgNO_3$) and characterized amplified alleles by their PCR product size.

To identify informative chloroplast markers, we selected five most polymorphic intergenic regions based on the chloroplast genome of *Catharanthus roseus* L. (Apocynaceae, Rauvolfioideae) (Ku *et al*. 2013 – GenBank accession NC_021423), a closely related species (Simões *et al*. 2016) (Table 2).

Table 2: Selected intergenic plastidial (cpDNA) regions amplified with product sizes based on *Catharanthus roseus* L. chloroplast genome (Ku *et al*. 2013).

Region	Primer sequence	Size (bp)
ndhD/psaC	*F*: 5'-CCGCAAATATTGGCAAAACTACAA-3'	129
	R: 5'-GCTAAACAAATAGCTTCTGCTCC-3'	129
trnH-GUG/psbA	*F*: 5'-CGCGCATGGTGGATTCACAATCC-3'	404
	R: 5'-GTTATGCATGAACGTAATGCTC-3'	404
trnR-UCU/atpA	*F*: 5'-GTCTAATGGATAGGACAGAGG-3'	115
	R: 5'-CCCTTTTGAAAGAAGCTATTCAGG-3'	115
ccsA/ndhD	*F*: 5'-TACTTCCGGGCTTTTAACTCA-3'	239
	R: 5'-GTACCCGGATTTCGTTCTTTC-3'	239
ndhG/ndhI	*F*: 5'-CCCAACAGTCCCAAGATTAGA-3'	239
	R: 5'-TCCCCGTTGTAGATTGGAAAT-3'	402

First, we conducted a pilot study considering eight samples of *R. weddelliana* and the following PCR protocol: 30 cycles of 1 min at 95°C, 1 min at 62°C and 1 min at 72°. We sequenced forward and reverse products using Macrogen sequencing facility (Macrogen Inc., Seoul, South Korea) and identified two polymorphic markers: *trnH*-GUG/*psbA* and *trnR*-UCU/*atpA*. These markers were then amplified and sequenced for all 77 samples.

Genetic diversity and population structure

To understand genetic differentiation within and among populations we first calculated general population metrics: allelic richness (A_R), observed heterozygosity (H_O), expected heterozygosity (H_E), inbreeding coefficient (F_{IS}) and population genetic differentiation (pairwise G'_{ST}, Hendrick 2005), using R packages "diveRsity" (Keenan *et al*. 2013). We used the same package to perform an exact test for Hardy-Weinberg Equilibrium, with 2000 replicates for the Monte Carlo procedure and Bonferroni correction (Rice 1989). To test for isolation by distance (IBD), we performed a Mantel test by comparing correlations for 1000 permutations of population's genetic distance (Nei's distance) with geographic distance using R packages "ade4" (Chessel *et al*. 2004) and "adegenet" (Jombart 2008).

To estimate population genetic structure, we performed a Bayesian clustering analysis in Structure 2.3.4 (Pritchard *et al*. 2000). We considered ten runs for each possible k-value, ranging from one to the total number of populations, with 5×10^5 burn-in period and 10^6 Markov Chain Monte Carlo (MCMC) repetitions. We processed results with the online tool CLUMPAK (Kopelman *et al*. 2015), which summarizes and plot results for multiple k-values using the algorithms DISTRUCT (Rosenberg 2004) and CLUMPP (Jakobsson & Rosenberg

2007). We also used CLUMPAK to estimate Evanno's Δk (Evanno *et al.* 2005), in order to detect the value of k that captures the uppermost level of structuration within our samples.

Additionally, we calculated mutation-scaled effective population sizes and mutation-scaled immigration rates using Migrate-n version 3.6.11 (Beerli & Felsenstein 2001; Beerli 2009) to inferred historical migration scenarios among the sampled localities. For that, we selected random subsamples of five individuals per population to overcome limitations of difference among our samples sizes and scaled migration rate estimates using geographical distance among populations, to consider potential environmental barriers between sites.

Phylogenetic analysis and divergence times

We evaluated, trimmed and generated consensus sequences for *trnH*-GUG/psbA and *trnR*-UCU/atpA contigs on Geneious 9.1.2 (Biomatters Ltd., Auckland, New Zealand). We aligned consensus sequences using the Geneious alignment algorithm and then manually checked the output. To identify haplotypes and phylogenetic groups, we generated a minimum spanning network with PopART software (Leigh & Bryant 2015) using a concatenated alignment for both markers.

We used Bayesian inference to describe the relationship among the haplotypes. We analysed alignments separately in jModelTest 2.1.10 (Guindon & Gascuel 2003; Darriba *et al*. 2012) to select the best-fitting substitution model for each marker, based on Akaike Information Criterion (Akaike 1974) corrected for small sample sizes (AICc - Hurvich & Tsai 1989). We performed the analyses for concatenated alignments using MrBayes 3.2.6 (Ronquist *et al*. 2012). We ran 10^6 generations, with two runs, four chains and trees sampling every 500 generations.

We estimated evolutionary rates and divergence times in BEAST2 2.4.7 (Bouckaert *et al*. 2014). We tested strict and relaxed clocks, with a substitution rate between $0.8 \pm 0.04 \times 10^{-9}$ and $1.52 \pm 0.06 \times 10^{-9}$ substitutions per site per year, as estimated for intergenic plastidial indels and chloroplast DNA, respectively (Yamane *et al.* 2006). Clock models were compared based on AICM values (Baele *et al*. 2012), calculated on the software Tracer 1.6.0 (Rambaut *et al*. 2014). To test demographic expansion for each major clade recovered by Bayesian inference phylogeny, we calculated Tajima's D (Tajima 1989) and Ramos-Onsins & Rozas R_2 (Ramos-Onsins & Rozas 2002) metrics with R package "pegas" (Paradis 2010), and applied a Coalescent Bayesian Skyline approach (Drummond *et al.* 2005) available in BEAST2. We processed resulting trees to generate a consensus tree in TreeAnotator (Bouckaert *et al*. 2014) and prepared tree graphics in FigTree 1.4.2 (Rambaut 2014).

Demographic scenarios simulation

We implemented an Approximate Bayesian Computation approach in order to test nine possible scenarios that account for the demographic history of the *R. weddelliana* clade according to the clusters recovered by the Structure analysis (Figure 1): (1) stability scenario (no changes in size and no dispersal/extinction); (2) LGM retraction scenario (effective size of all populations reduced during LGM and gradually expanded on mid-Holocene); (3) mid-Holocene retraction (effective size of all populations reduced during mid-Holocene and gradually expanded to present); (4) northward colonization scenario (populations from south gradually expanded and colonized north); (5) southward colonization scenario (populations from north gradually expanded and colonized south); (6) peripheral-core colonization scenario (marginal populations colonized core regions during LGM); (7) core-peripheral

colonization scenario (core populations colonized peripheral regions during LGM); (8) LGM refugia scenario (populations from non-refugia areas went extinct during LGM and were latter colonized by populations from Chapada dos Guimarães and Goiás on mid-Holocene); and (9) interglacial refugia scenario (populations 1,3, and 5 went extinct during LIG and were colonized by populations 2 and 4 from on LGM).

For coalescent simulations, we adopted one year as generation time, based on reproductive maturity reported for *Rauvolfia serpentina* (L.) Benth. ex Kurz (Dutta *et al*. 1963). We adopted 8.87×10^{-4} mutations per site per year as an estimate of mutation rate for SSR markers, based on *Arabidopsis thaliana* SSR mutation rates (Marriage *et al*. 2009). We performed data simulations (100,000 for each model) with the scripts given in Perez *et al* (2016), using empirical sample sizes. We implemented θ values calculated by Migrate-n as an initial approximation of the estimate effective population sizes. We established priors for the parameters from simulated uniform distributions using a distribution of 115000–130000 (LIG), 16000–31000 (LGM), and 5000–7000 (mid Holocene) years for divergence time (τ) and a distribution of 50 to 10,000 individuals for the effective population size of the ancient population (N_e). For estimates of rates of long dispersal (used in the models 4 to 9) we computed the contraction rates during colonization (θrF-A) as the ratio between the θ value of the ancient population (also drawn from an uniform distribution ranging from 0.001 to 0.1) and the magnitude of population expansion after colonization (θrC-A), which we estimated as the ratio between the current θ and the θ value of the ancient population (sampled from 0.01 to 1).

We converted the nuclear coalescent genealogies data obtained by our simulations into microsatellite data using a single step mutation model (SMM) available in the software

"microsat" (Cox 2011; available at: http://massey.genomicus.com/software.html#microsat). Then, we converted the SSR output using a script developed by Perez *et al*. (2016) and used the software Arlequin version 3.5 (Excoffier & Lischer 2010) to calculate the number of alleles (A), the expected heterozygosity (H_E), and a modified version of Garza and Williamson's M (mM) (Excoffier *et al*. 2007; Garza & Williamson 2001). For the empirical and simulated cpDNA sequence data, we calculated the proportion of polymorphic sites (π), the number of segregating sites (S), Tajima's D (Tajima 1989) and θH (Fay & Wu 2000) parameters using a custom PERL script developed by N. Takebayashi (available at: http://raven.iab.alaska.edu/~ntakebay/teaching/programming/coalsim/scripts/msSS.pl). We employed the empirical sets of summary statistics for running the ABC modeling, using the multinomial logistic regression and the neural network methods available in the R package "abc" version 1.4 (Csilléry *et al*. 2012). We adopted a threshold level of 0.01 and obtained a total of 9,000 simulations, which were retained in the posterior probabilities.

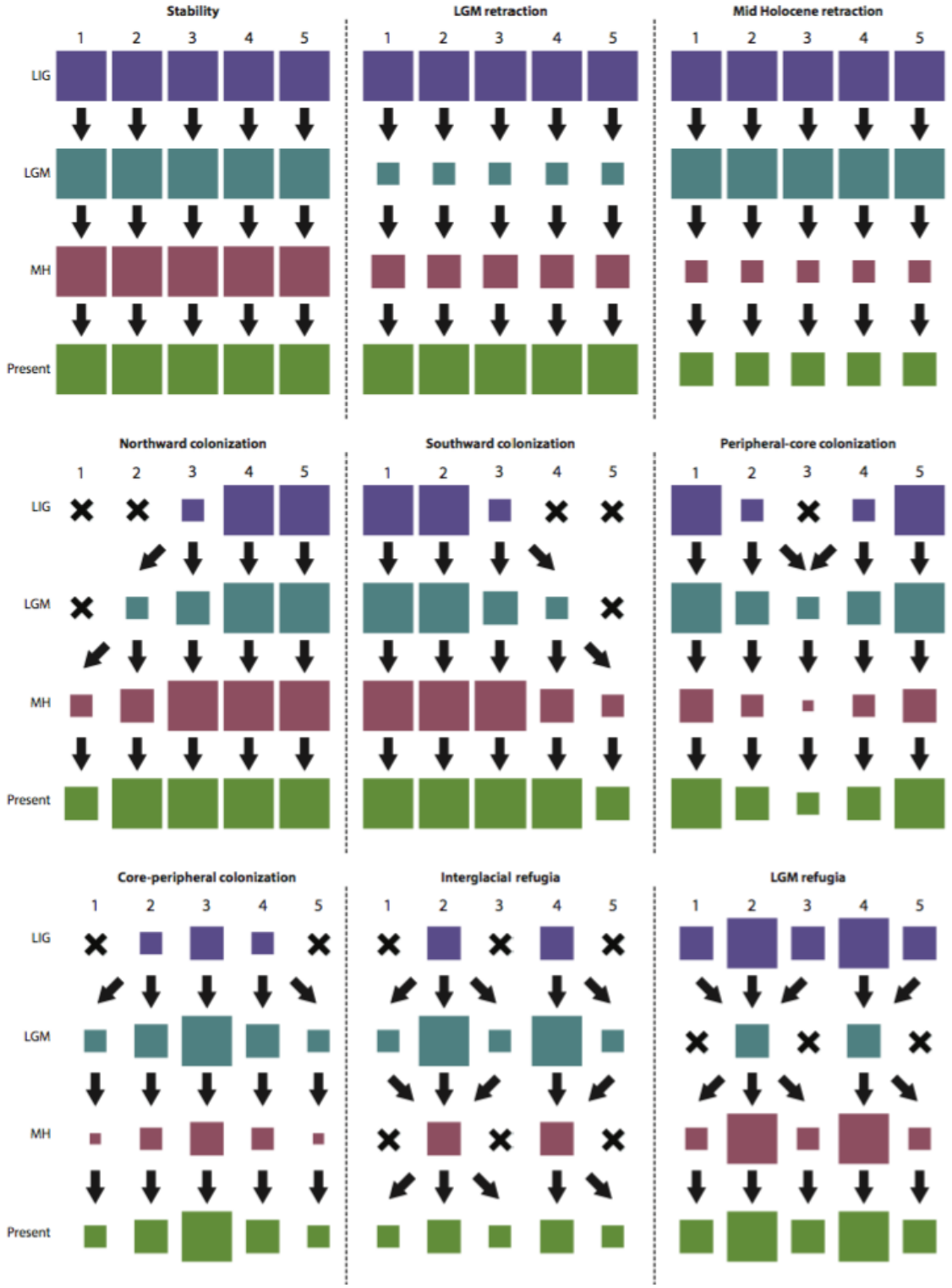


Figure 2: *Rauvolfia weddelliana* clade demographic scenarios tested with ABC. Boxes represent putative populations inferred by STRUCTURE analysis. *Stability*: No changes in size and no dispersal/extinction; *LGM retraction*: All populations reducing effective size during LGM and gradually expanding on MH; *Mid Holocene retraction*: All populations reducing effective size during MH and gradually expanding on present; *Southward colonization*: populations from north gradually expanding and colonizing south. *Northward colonization*: populations from south gradually expanding and colonizing north; *LGM refugia*: populations 1,3, and 5 gone extinct during LGM and being colonized by populations 2 and 4 on MH; *Interglacial refugia*: populations 1,3, and 5 gone extinct during LIG and being colonized by populations 2 and 4 on LGM.

Species distribution modelling

For the development of potential distribution models, we retrieved distribution data from online collections, herbarium surveys and previous studies (Koch 2002). We manually verified specimen identification and geographic localities' information. We listed 71 unique occurrences for *R. weddelliana*, and less than five ones for *R. gracilis*. Since we were unable to recover sufficient points for generating accurate models, we modelled the potential distribution of *R. weddelliana* (narrow scenario) and *R. gracilis* combined with *R. weddelliana* (broad scenario). This approach is based on the premise of niche conservatism over closely related species (Peterson *et al*. 2011) and has been used as an alternative to estimate niches to species with restricted distributions and small number of samples (Bonatelli *et al*. 2014). To test the hypothesis of niche equivalency, which would support the broad scenario approach, we used statistics of niche identity Warren's I (Warren *et al*. 2008) and Schoener's D (Schoener 1968), with 1000 bootstrap replicates.

Using the set of verified occurrences explained above, we developed geographic buffers (*i.e*. 2-degrees of radius circles) around points using Q-GIS (QGIS Development Team, 2016) vector processing tools. We extracted background data for all 19 bioclimatic variables (30-sec resolution) provided by WorldClim dataset (Hijmans *et al*. 2005) by using a random set of 1000 points generated inside the buffer. We calculated pair-wise Pearson's correlation for all environmental layers and randomly excluded one of the variables from highly correlated (>0.7) pairs (Dormann *et al*. 2013). We applied this step to reduce the bias of incorporating redundant variables in the models. We obtained a subset of 13 less correlated variables, which included "*bio 1*" (Annual Mean Temperature), "*bio 2*" (Mean Diurnal

Range), "*bio 3*" (Isothermality), "*bio 4*" (Temperature Seasonality), "*bio 5*" (Max Temperature of Warmest Month), "*bio 7*" (Temperature Annual Range), "*bio 10*" (Mean Temperature of Warmest Quarter),"*bio 13*" (Precipitation of Wettest Month),"*bio 14*" (Precipitation of Driest Month),"*bio 15*" (Precipitation Seasonality), "*bio 18*" (Precipitation of Warmest Quarter), "*bio 19*" (Precipitation of Coldest Quarter) and elevation. We used MAXENT (Phillips *et al*. 2006) to generate distribution models for present scenario and projected to CCSM4 and MIROC reconstructions of Mid-Holocene (ca. 6,000 years before present), Last Glacial Maximum (ca. 21,000 years before present) and Last Interglacial (ca. 120,000 years before present). We adopted default MAXENT values for the extant parameters. We converted generated models to presence/absence maps based on the lowest suitability value predicted for actual occurrence points (Peterson *et al*. 2011) and summed generated occurrence maps with Q-GIS raster calculator to obtain stability maps. Populations were considered stable if predicted as presence areas (i.e. suitability > threshold) for all four climatic scenarios.

Results

Genetic diversity and population structure

Allelic richness and gene diversity were higher for central populations VN (A_R = 3.86, H_E = 0.68) and CP (A_R = 3.84, H_E = 0.66), while the central population SV presented the lower nSSR genetic diversity (A_R = 2.73, H_E = 0.52) (Table 1). Population SV was also the only population with significant deviation from Hardy-Weinberg Equilibrium (*p* value = 0.021).

Bayesian structure analysis of microsatellite markers revealed five genetic clusters (K = 5) (Figure 4). Of these, four clusters are geographically structured, being predominantly associated with central (clusters 2 and 3) and southern populations (clusters 4 and 5). Cluster 1 presented a disjoint geographical distribution, with genotypes present on northern VL populations and central SV and RP populations. We found no correlation between genetic and geographic distances (Mantel test *p*-value = 0.586). We also inferred higher migration from RP (central) to CD (southern) populations ($M_{CD \rightarrow RP}$ = 662.31), from CD (southern) to VN (central) ($M_{CD \rightarrow VN}$ = 622.04), from PA (southern) to CD (southern) ($M_{PA \rightarrow CD}$ = 613.30), and from VL (northern) to CP (central) ($M_{VL \rightarrow CP}$ = 622.04) (Figure 5; Supplementary data, Table 1).

Genetic differentiation measure (G'_{ST}; Hendrick 2005) values ranged from 0.1218 between central populations VN and CP to 0.7594 between central populations RP and SV (Table 3). Higher G'_{ST} values were mostly concentrated between central populations SV and RP in relation with southern populations CD and PA (SV/CD = 0.7024; SV/PA = 0.6044; RP/CD = 0.7191; RP/PA = 0.6917). The lowest values were identified between populations from the same geographical region for central populations CP and VN (G'_{ST} = 0.1218) and south populations CD and PA (G'_{ST} = 0.3064). We found intermediate values of G'_{ST} (0.4055–0.6395) for northern population (VL), composed by *R. gracilis*.

Table 3: Pairwise G'_{ST} values (Hendrick, 2005) for *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I. Koch & Kin.-Gouv. populations.

		North		Central			South	
Region	Population	VL	CP	VN	SV	RP	CD	PA
North	VL	-						
Central	CP	0.417	-					
	VN	0.4923	0.1218	-				
	SV	0.6395	0.4643	0.4333	-			
	RP	0.5098	0.5912	0.641	0.7594	-		
South	CD	0.4055	0.5893	0.5686	0.7024	0.6917	-	
	PA	0.4114	0.4358	0.3698	0.6044	0.7191	0.306	-

We obtained 622 and 128 base pairs sequences for *trnH-psbA* and *trnR-atpA*, respectively. We obtained 25 segregating sites and 16 parsimony informative sites from the concatenated sequence alignment. The most suitable substitution model for *trnH*-GUG/*psbA* and *trnR*-UCU/atpA regions were HKY85 (Hasegawa *et al*. 1985) and F81 (Felsenstein 1981), respectively. Bayesian inference phylogeny recovered two major groups, one composed by southeast populations (PA and CD) and other composed by central and north populations (Figure 2). *Rauvolfia gracilis* formed a polytomy with *R. weddelliana* individuals from the northwest clade (SV, RP, and VN). Populations from Chapada dos Guimarães (VN and CP) were recovered as a paraphyletic group, with three distinct clades within northwest clade.

Chloroplast sequences haplotype network also supported the northwest-southeast division for *Rauvolfia weddelliana* clade (Figure 3). The most abundant haplotype was haplotype I, followed by haplotypes VII and V, respectively. Despite being the more abundant, haplotype I was restricted to Chapada dos Guimarães populations (VN and CP), along with other three exclusive haplotypes (II, IV and VI). The most widespread haplotype

was haplotype V, present in three populations (SV, RP, and VL). Although the *R. gracilis* population presented a unique haplotype (haplotype III), most of its individuals shared haplotype V with populations from south of Mato Grosso (SV and RP). Southern populations (PA and CD) presented two considerably differentiated haplotypes (haplotypes VII and VIII, with 11 mutation steps), which are not present in northwest group (Figure 3).

Phylogenetic analysis and divergence times

Bayesian phylogenetic tree recovered three major clades for *R. weddelliana* and *R. gracilis* populations (Figure 2). The southeast clade included individuals from populations CD and PA, while the two northwest clades presented one clade composed exclusively by individuals from central populations CP and VN, and another one including individuals from all five northwest populations (CP, VN, VL, RP and SV).

A strict clock model was discarded when compared to relaxed lognormal clock model, based on AICM values ($AICM_{relaxed} > AICM_{strict}$). The divergence of most lineages divergences was estimated to the late Miocene. The major split between southeastern and northwestern lineages took place at 10.5 mya. The northern clade lineages also diverged on late Miocene, at 8.3 mya. Tajima's D and R_2 neutrality tests for northern clades I were not significant (p-value > 0.1). For southern clade, Tajima's D parameter was statistical significant (p-value < 0.01) and presented a negative value (clade III Tajima's D = -2.11). We could not calculate any of these statistics to clade II, since no segregating sites were found between sequences of this clade. We could not discard demographic stability with the Bayesian Coalescent Skyline for all clades, due to the high posterior density intervals associated with plots (Supplementary data, Figure 1).

Demographic scenarios simulation

The posterior probabilities of ABC simulations indicate the interglacial refugia, with intermediate populations going extinct during LIG and being latter colonized by refugia populations on LGM (Model 8) as the most likely scenario ($PP_{neuralnet}$ = 0.659; $PP_{mnlogist}$ = 0.9998; Table 4). All models assuming retractions during glacial climates (Models 2 and 9) presented poor performances, with posterior probabilities lower than 0.007. The stability scenario (Model 1) presented the lower posterior probability ($PP_{neuralnet}$ = 0.0029; $PP_{mnlogist}$ = 0).

Table 4: Demographic models tested for *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I.Koch & Kin.-Gouv., with their respective posterior probabilities estimated with neural network and multinomial logistic regression methods.

Model	*Description*	$PP_{neuralnet}$	$PP_{mnlogistic}$
1	Stability	0.0029	0
2	LGM retraction	0.0047	0
3	Mid Holocene retraction	0.0044	0
4	Northward colonization	0.1524	0.0002
5	Southward colonization	0.1051	0
6	Peripheral-core colonization	0.0614	0
7	Core-peripheral colonization	0.0034	0
8	Interglacial refugia	0.6590	0.9998
9	LGM refugia	0.0068	0

Species distribution modelling

Narrow and broad scenarios showed important differences on range reconstruction for *R. gracilis* and *R. weddelliana*, and both presented high AUC values (narrow scenario AUC =

0.967; broad scenario AUC = 0.926). Niche identity tests (Warren's I = 0.502; Schoener's D = 0.301), however, did not support the equivalency of *R. weddelliana* and *R. gracilis* niches (p-value < 0.01), indicating that narrow scenario is more suitable to our data (Supplementary data, Figure 2 and 3). Thus, we based our interpretations of geographic distribution of the species solely on *R. weddelliana* models. For this scenario, our models indicated a surprisingly pattern of expansion during Holocene, followed by a slight retraction between 6,000 ypb and the present distribution of *R. weddelliana* (Figure 6). The geographical distribution of species already displayed a clear north/south subdivision during Last Interglacial (LIG, 120,000 ybp). Latter, the cold and dry climatic conditions of Last Glacial Maximum promoted a strong retraction in the group distribution, which restricted suitable areas to central regions of the distribution (see Chapada dos Guimarães, circle 2 in Figure 6). Climatic conditions of mid Holocene considerably enhanced the potential range for the species, and models recovered a strong expansion on potential distribution of species during this period. *Rauvolfia gracilis* population was recovered as a relictual distribution of this widespread northern distribution of *R. weddelliana* during mid-Holocene (Figure 6).

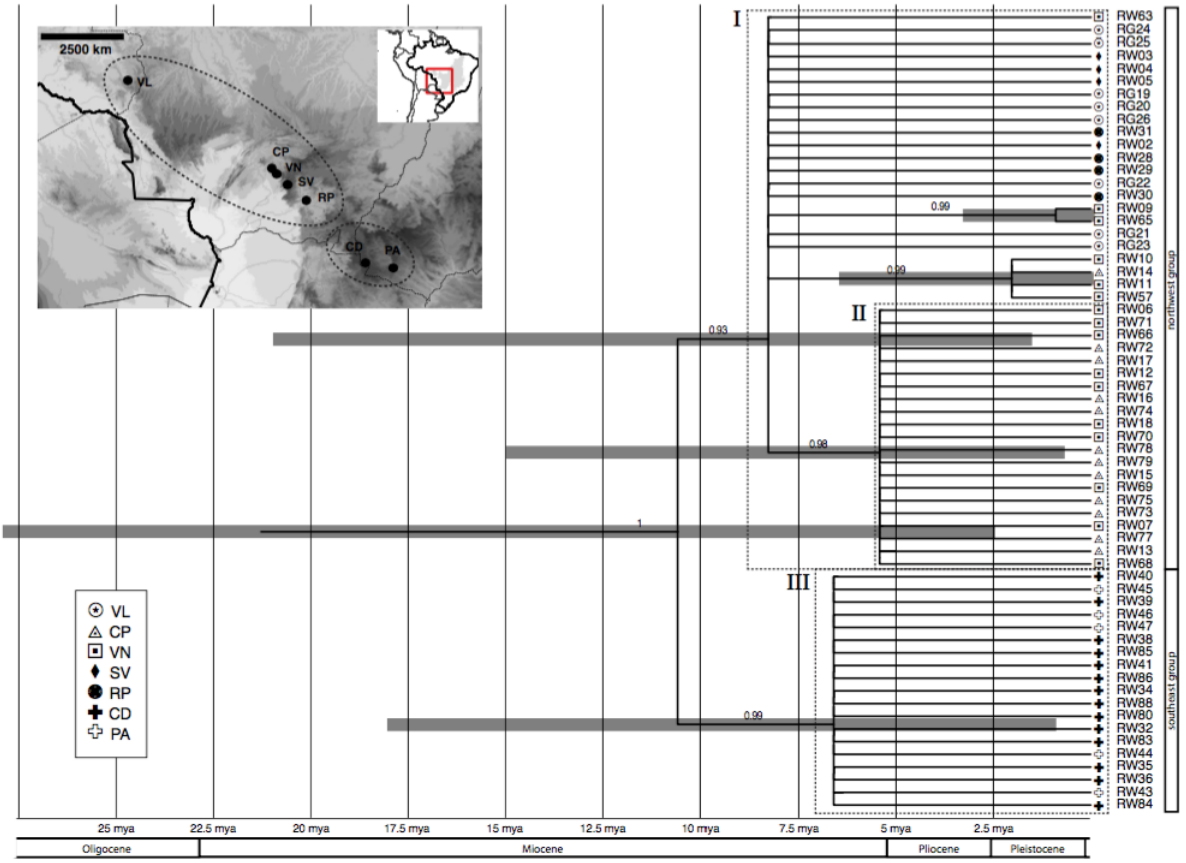


Figure 3: Dated Bayesian inference phylogenetic tree for *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I. Koch & Kin.-Gouv. samples from seven populations, based on concatenated chloroplast intergenic spacers *trnR*-UCU/*atpA* and *trnH*-GUG/*psbA*, with 3 major clades and two geographical groups Tip symbols represent sample geographic population. Numbers above each branch represents branch supports and branch lengths are proportional to genetic distance. Bars represent node age intervals (height 95% HDP). On the upper left figure, an elevation map (darker tones indicating higher elevation) with the location of populations and groups.

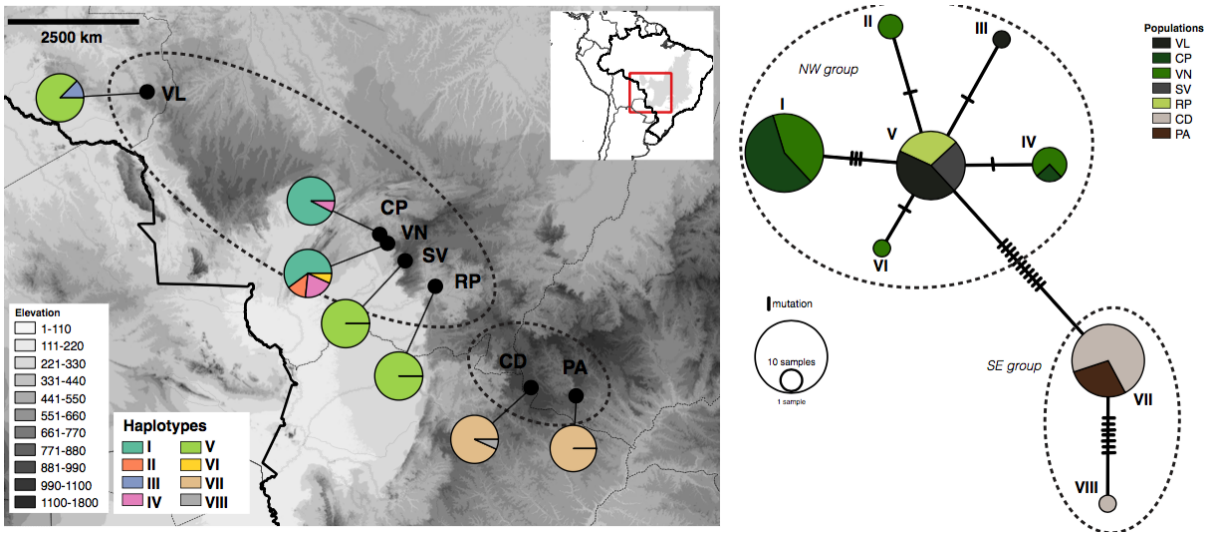


Figure 4: Geographic distribution of cpDNA haplotypes for *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I. Koch & Kin.-Gouv. in Central-western Brazil. (A) Elevation map for central-western Brazil indicating *R. weddelliana* and *R. gracilis* population geographic distributions. (B) Minimum spanning network for concatenated cpDNA markers. Each circle represents a haplotype (I, II, III, IV, V, VI, VII, VIII) and the circles radius is proportional to haplotype frequency. Color fills represent populations and line dashes correspond to mutations.

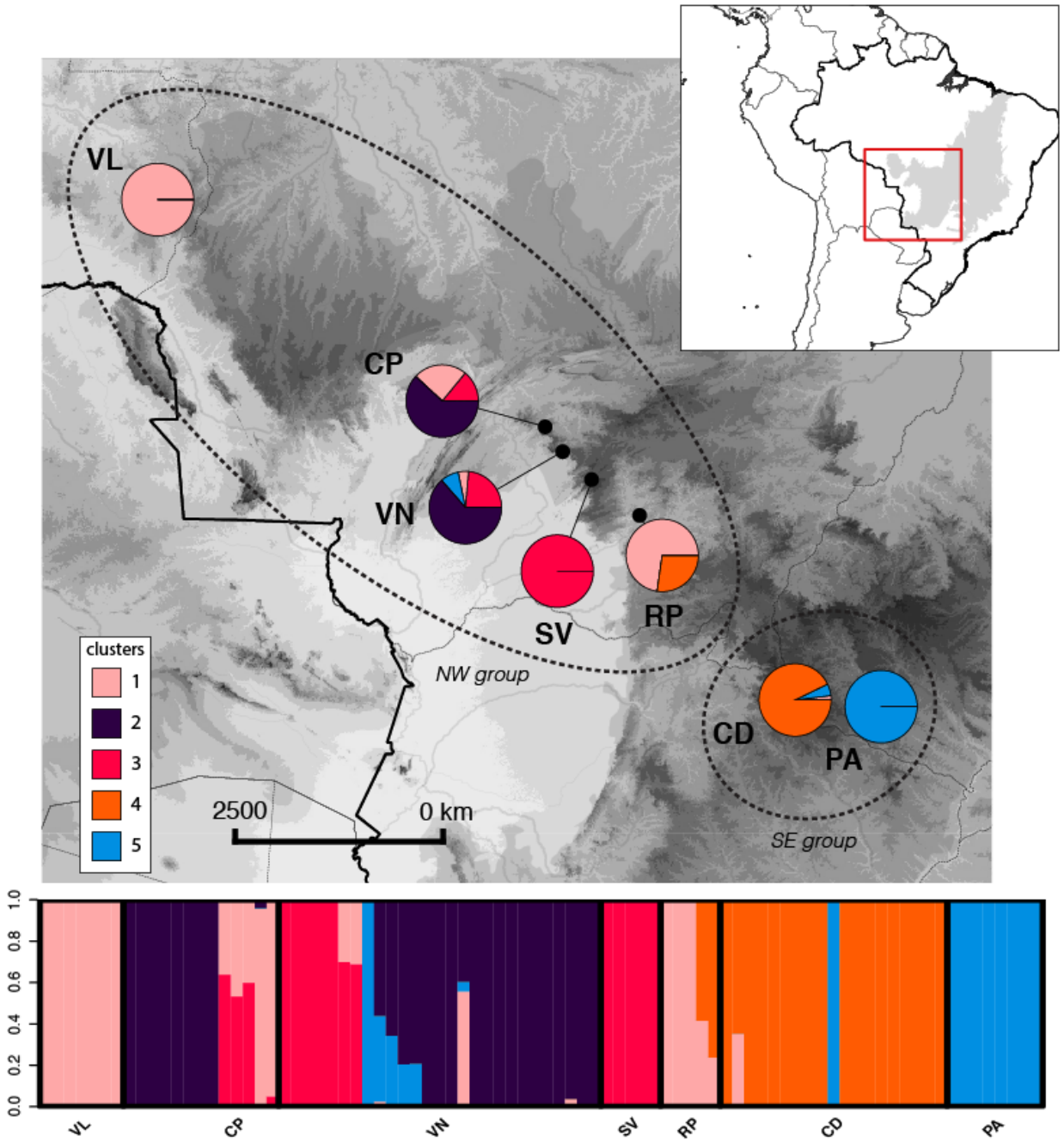


Figure 5: Population structure for *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I. Koch & Kin.-Gouv. inferred by Structure (Pritchard *et al*. 2000). Pie charts represent the proportion of individuals from each of the five clusters occurring in each geographic population, and vertical bars correspond to the individual membership probability of each sample. Map background represents elevation, with darker tones indicating higher areas.

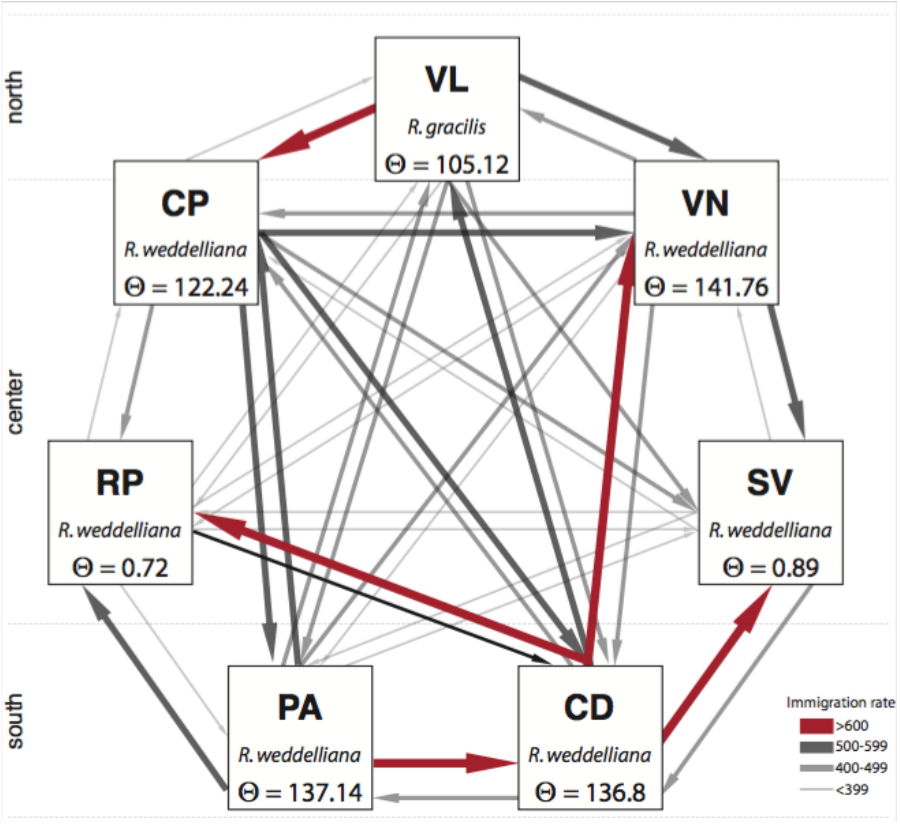


Figure 6: *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I. Koch & Kin.-Gouv. mutation scaled immigrants for north (VL), center (CP, VN, RP, CV), and south (PA, CD) populations. *Θ*: mutation scaled effective population sizes. Arrows width and color proportional to immigration rate.

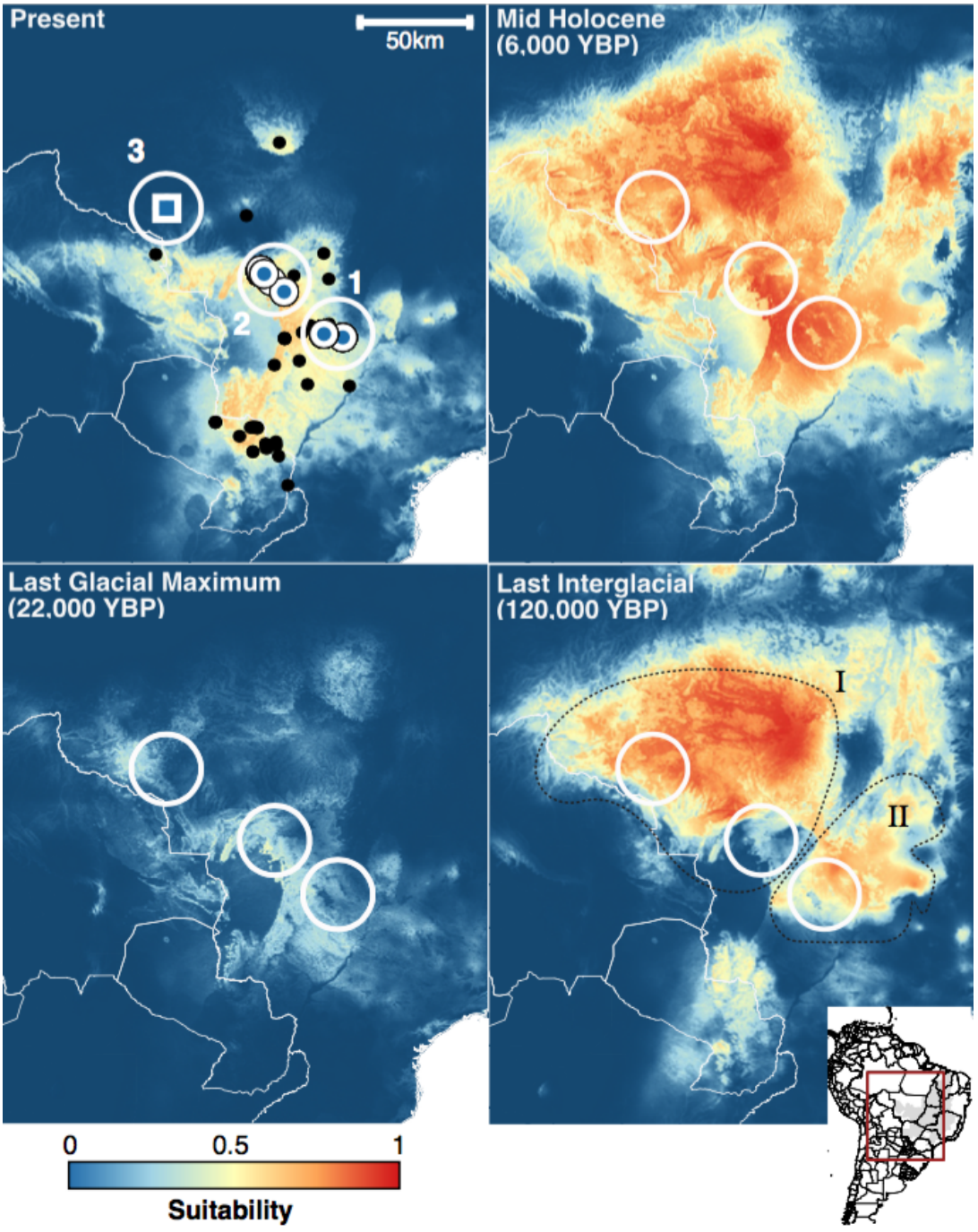


Figure 7: MAXENT models for *Rauvolfia weddelliana* Müll.Arg. potential distribution for present, mid-Holocene (6,000 years before present) and Last Glacial Maximum (21,000 years before present). White and blue colored circles and square indicate sampled localities for *R. gracilis* and *R. weddelliana*, respectively. Black circles represent database occurrence points for *R. gracilis* and *R. weddelliana*, respectively, used for generating the distribution models. Colder colors indicate low suitability, while hot colors indicate high values of suitability.

Discussion

In this study we describe genetic diversity and phylogeographic structure and modelled the distribution of *R. gracilis* and *R. weddelliana* to understand the impacts of climate in the range shifts within this group since the Last Interglacial (LIG: 120,000 years before present). We also evaluated the suitability of different demographic scenarios that may have shaped the current distribution of the lineages.

Our results suggest that geomorphological processes from the Pliocene-Pleistocene boundary, along with interglacial refugia dynamics, shaped the diversification of the *R. weddelliana* complex. Our results point to a north-south division of the group, which predated the Last Glacial Maximum. We attribute the lack of a signature of glacial expansion is related to limitations of niche modelling approaches to incorporate important edaphic variables in past climatic scenarios. Demographic simulations recovered a scenario of interglacial extinction glacial recolonization of populations within *R. weddelliana* group, typical of interglacial refugia scenarios. Thus, these ancient diverging lineages occurring in northwestern and southeastern regions underwent range expansion in last glaciations and subsequent retractions during interglacials, generating the current distribution of the *R. weddelliana* group.

Also, our data do not support the current taxonomic division between *R. gracilis* and *R. weddelliana*, indicating that these species might not be reproductively isolated and therefore should not be considered as different species. Yet, we identified a cryptic, highly differentiated lineage occurring on the southern distribution of *R. weddelliana*, and further morphological studies are recommended to test the degree of differentiation of these populations and its taxonomic interpretation.

Relative impact of geomorphological processes and glaciations

We detected a strong geographic structure between northern and southern distribution of *R. weddelliana* and *R. gracilis*. The split of *R. weddelliana* northwest and southeast lineages took place in late Miocene, which is consistent with the uplift of the Central Brazilian Plateau (Silva 1997), suggesting that geographic processes prior to Pleistocene climatic shifts shaped the divergence of these major lineages. The genetic structure observed for both markers is consistent with an allopatric differentiation between northern and southern populations, with subsequent secondary contact. This north-south divergence likely occurred previously to the Pleistocene period, despite low statistical support for the node ages recovered in phylogenetic analyses. Our results are consistent with late Pliocene diversification ages already proposed by the few studies available for other taxa occurring in western Cerrado plateaus, indicating that this might be a common pattern. Studies for other species occurring on plateaus, like the lizard *Micrablepharus atticulolus* and the frog *Hypsiboas albopunctatus*, also revealed divergence ages prior to Pleistocene (Prado *et al*. 2012; Santos *et al*. 2014). Further studies with other taxa may help clarifying the extent of the impacts of geomorphological processes in the diversification of Western Cerrado biota. So far, it seems to be a primary process shaping lineages on the region.

The low geographic structure within the northwestern group suggests higher levels of gene flow for populations in this area compared to the south populations. This pattern may be interpreted as a lower relative effectiveness of barriers separating populations within the northwestern group. For the asymmetric immigration rates from southeast to northwest, the simplest explanation is that the pool of potential dispersers in the first region is greater than in

the second (Kaweki & Holt 2002). Other possible explanations are the relative probability of dispersal between populations, which may be affected by physical processes of transportation of the propagules (e.g. wind, river systems), or even by differences in habitat quality (Kaweki & Holt 2002), since most of northwestern populations were outside conservation units and thus, facing considerable anthropization.

Reproductive traits and mating systems can also affect the population structure. The mating system in the *R. weddelliana* complex are still not described, but other species of the genus present functional dioecy and mechanisms like self-incompatibility and herkogamy to prevent self-fertilization (Lopes & Machado 1999; Koch *et al*. 2002), which may differently affect the spatial genetic structure of populations (Clegg 1980; Loveless & Hamrick, 1984). The low population density and the patchy distribution of *R. weddelliana* and *R. gracilis* are expected to decrease pollinators visits, which can also negatively impact effective population size (N_e) and gene flow, while increases inbreeding (Loveless & Hamrick 1984). Also, species discontinuous distribution can lead to genetic isolation, which is one major factor associated with increasing rates of inbreeding, especially for small populations like the ones of *R. weddelliana*, and consequent loss of genetic diversity through genetic drift (Ellstrand 1992; Ellstrand & Elam 1993). The pollination in *Rauvolfia* is most likely mellitophylous (bees) and psychophilous (butterflies) (Lopes & Machado 1999, Kimmel *et al*. 2010), and fruit-eating birds disperse seeds (Rao 1956; Snow 1981; Kimmel *et al*. 2010). Animal-ingested dispersal system observed in *Rauvolfia* is expected to reduce genetic structure between populations, since seeds are regularly long-distance dispersed (Loveless & Hamrick 1984). We consider these dispersal mechanisms to have played an important role in the maintenance of gene flow

on the observed geographic structure for *R. weddelliana* and *R. gracilis*, as we found no isolation by distance (IBD) pattern (Mantel test *p*-value = 0.586).

Demographic history of populations

Although we were not able to detect climatic induced allopatry through the species distribution models, *R. weddelliana* and *R. gracilis* demographic simulations recovered strong evidences of typical interglacial refugia processes on the populations of the group. Performances of the modeled stable demography (Model 1) and glacial retraction scenarios (Models 2 and 9) were rejected when compared to a glacial expansion scenario (Model 8). This result suggests that the genetic structure and coalescence times of the present nuclear and chloroplast lineages are consistent with a scenario where lineages from stable areas, on the center and southern regions, underwent range expansions during last glacial maximum and subsequently retracted their geographic ranges during interglacial periods.

Controversially, *R. weddelliana* species distribution models recovered a glacial retraction scenario (Figure 1, Model 9), which was not supported by our coalescent simulations. This result was surprising, although not unique, since other distribution model already reported glacial retraction scenarios within Cerrado (Werneck *et al*. 2012). Nonetheless, other lines of evidences, like pollen (van der Hammen 1974) and soil deposition profiles (Sanaiotti *et al*. 2002), do reject the strong range retraction for the arid Cerrado vegetation, as depicted by our modeled distribution. Therefore, our interpretation is that the niche model we developed was probably biased due to the restricted and patchy distribution of the species. Also, potential limitations are expected due to the incapacity of past models to

incorporate edaphic components, which are undoubtedly one of the most important elements determining the distribution of plants and significantly affect the performance of niche models (Bertrand *et al*. 2012). Although the combination of niche models with demographic simulations is recommended, distribution models are limited in the aspect that they predict a potential area of suitable environmental conditions, not necessarily a demographic distribution, which can generate incongruences between these evidences (Dellicour *et al*. 2014).

Despite being widely discussed in literature, there is still controversy about the relative role of the Pleistocene interglacial refugia driven diversification in South American biogeography, when compared with more ancient processes and in many cases observed patterns contradict theory's predictions for different groups of organisms. Distribution models and phylogenetic structure for several other Cerrado species also support range retractions during the interglacials of Holocene (Collevatti *et al*. 2012; Bonatelli *et al*. 2014). Our results reinforce this pattern of response for arid species, which coupled with the tectonic processes of the Miocene, shaped the current distribution of lineages of *R. weddelliana* and *R. gracilis*.

Conclusions

We detected a strong geographic structure between northern and southern distribution of *R. weddelliana* and *R. gracilis*. All evidences support a prior to Pleistocene divergence between these lineages. Within north and southern clades, ABC simulations and niche models suggested contrasting scenarios. While coalescent simulations indicate recent interglacial retraction, niche models suggested glacial retractions. We argue that the scenario proposed by niche models is less precise than the demographic simulations, especially due to the absence

of soil predictors in past climatic scenarios used in the niche models. Thus, our interpretation is that the current distribution of genetic diversity and phylogeographic structure within north and south groups of *R. weddelliana* and *R. gracilis* represents an interglacial refugia scenario. Additionally, due to the ancient divergence time estimated by the relaxed molecular clock for the split between northern and southern lineages, geomorphological events from late Miocene may have also played an important role on shaping the lineages modern distribution of *R. weddelliana*.

References

Ab'Sáber A.N. (1977) O domínio morfoclimáticos da América do Sul. Primeira aproximação. *Geomorfologia*, **52**, 1–22.

Ab'Sáber A.N. (1983) O domínio dos cerrados: introdução ao conhecimento. *Revista do Serviço Público*, **111**, 41–55.

Aitken S.N., Yeaman S., Holliday J.A., Wang T., & Curtis-McLane S. (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Akaike H. (1974) A new look at the statistical model identification. System identification and time-series analysis. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Baele G., Lemey P., Bedford T., Rambaut A., Suchard M.A., & Alekseyenko A. V. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. *Molecular Biology and Evolution*, **29**, 2157–2167.

Barbosa A.C. de O.F., Collevatti R.G., Chaves L.J., Guedes L.B.S., Diniz-Filho J.A.F., & Telles M.P. de C. (2015) Range-wide genetic differentiation of *Eugenia dysenterica* (Myrtaceae) populations in Brazilian Cerrado. *Biochemical Systematics and Ecology*, **59**, 288–296.

Beerli, P. (2009) How to use Migrate or why are Markov Chain Monte Carlo programs difficult to use? In G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi, editors, Population Genetics for Animal Conservation, volume 17 of Conservation Biology, pp. 42–79. Cambridge University Press, Cambridge UK, 2009.

Beerli, P. and J. Felsenstein (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the USA*, **98**, 4563–4568.

Bertrand R., Perez V., & Gégout J.C. (2012) Disregarding the edaphic dimension in species distribution models leads to the omission of crucial spatial information under climate change: The case of *Quercus pubescens* in France. *Global Change Biology*, **18**, 2648–2660.

Bonatelli I.A.S., Perez M.F., Peterson A.T., Taylor N.P., Zappi D.C., Machado M.C., Koch I., Pires A.H.C., & Moraes E.M. (2014) Interglacial microrefugia and diversification of a cactus species complex: Phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. *Molecular Ecology*, **23**, 3044–3063.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.H., Xie D., Suchard M.A., Rambaut A., & Drummond A.J. (2014) BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, **10**, 1–6.

Caetano S., Prado D., Pennington R.T., Beck S., Oliveira-Filho A., Spichiger R., & Naciri Y. (2008) The history of Seasonally Dry Tropical Forests in eastern South America: Inferences from the genetic structure of the tree *Astronium urundeuva* (Anacardiaceae). *Molecular Ecology*, **17**, 3147–3159.

Chessel D., Dufour A.B., & Thioulouse J. (2004) The ade4 package - I : One-table methods. *R News*, **4**, 5–10.

Clegg M.T. (1980) Measuring plant mating systems. *BioScience*, **30**, 814–818.

Collevatti R.G., Grattapaglia D., & Hay J.D. (2003) Evidences for multiple maternal lineages of *Caryocar brasiliense* populations in the Brazilian Cerrado based on the analysis of chloroplast DNA sequences and microsatellite haplotype variation. *Molecular Ecology*, **12**, 105–115.

Collevatti R.G., Rabelo S.G., & Vieira R.F. (2009) Phylogeography and disjunct distribution in *Lychnophora ericoides* (Asteraceae), an endangered cerrado shrub species. *Annals of Botany*, **104**, 655–64.

Collevatti R.G., de Castro T.G., Lima J. de S., & Telles M.P. de C. (2012) Phylogeography of *Tibouchina papyrus* (Pohl) Toledo (Melastomataceae), an endangered tree species from

rocky savannas, suggests bidirectional expansion due to climate cooling in the Pleistocene. *Ecology and Evolution*, **2**, 1024–1035.

Csilléry K., François O., Blum M.G.B. (2012) ABC: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution*, **3**, 475–479.

Darriba D., Taboada G.L., Doallo R., Posada D. (2012) JMODELTEST 2: more models, new heuristics and parallel computing. *Nature Methods*, **9**, 772.

Davis, M.B. & Shaw R.G. (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, **292**, 673–679.

Dellicour S., Fearnley S., Lombal A., Heidl S., Dahlhoff E.P., Rank N.E., & Mardulyn P. (2014) Inferring the past and present connectivity across the range of a North American leaf beetle: Combining ecological niche modeling and a geographically explicit model of coalescence. *Evolution*, **68**, 2371–2385.

Diniz-Filho J.A.F., Bini L.M., Oliveira G., Barreto B. de S., Silva M.M.F.P., Terrible L.C., Rangel T.F.L.V.B., Pinto M.P., Sousa N.P.R., Vieira L.C.G., Melo A.S., de Marco Júnior P., Vieira C.M., Balmires D., Bastos R.P., Carvalho P., Ferreira L.G., Telles M.P. de C., Rodrigues F.M., Silva D.M., Silva Júnior N.J., & Soares T.N. (2009) Macroecologia, biogeografia e áreas prioritárias para conservação no cerrado. *Oecologia Brasiliensis*, **13**, 470–497.

Dormann C.F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., Marquéz J.R.G., Gruber B., Lafourcade B., Leitão P.J., Münkemüller T., Mcclean C., Osborne P.E., Reineking B., Schröder B., Skidmore A.K., Zurell D., & Lautenbach S. (2013) Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 027–046.

Doyle J.J. & Doyle J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue, *Phytochemical Bulletin*, **19** (1), 11-15.

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185–1192.

Dutta P.K., Chopra I.C., & Kapoor L.D. (1963) Cultivation of *Rauvolfia serpentina* in India. *Economic Botany*, **17**, 243–251.

Eiten G. (1972). The Cerrado vegetation of central Brazil. *Botanical Review*, **38**, 201–341.

Ellstrand N.C. (1992) Gene flow by pollen: implications for plant conservation genetics. *Oikos*, **63**, 77–86.

Ellstrand N.C. & Elam D.R. (1993) Population Genetic Consequences of Small Population Size: Implications for Plant Conservation. *Annual Review of Ecology and Systematics*, **24**, 217–242.

Evanno G., Regnaut S. & Gouldet J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611-2620.

Excoffier, L., Laval, G., & Schneider, S. (2007). Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Excoffier, L. & Lischer, H.E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Fay, J. C., & Wu, C. I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, **155**, 1405–1413.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368–376.

Françoso R.D., Haidar R.F., & Machado R.B. (2016) Tree species of South America central savanna: endemism, marginal areas and the relationship with other biomes. *Acta Botanica Brasilica*, **30**, 78–86.

Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology*, **10**, 305–318.

Guindon S. & Gascuel O. (2003). A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology*, **52**, 696–704.

Haffer J. (1969) Speciation in Amazonian Forest Birds. *Science*, **165**, 131–137.

Hasegawa M., Kishino H., & Yano T. aki (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

Hedrick P.W. (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633.

Hewitt G.M. (1996) Some genetic consequences of ice ages, and their role, in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hewitt G.M. (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 183–195.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965-1978.

Hurvich C.M. & Tsai C.-L. (1989) Regression and time series model selection in small samples. *Biometrika*, **76**, 297–307.

Jakobsson M. & Rosenberg N.A. (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Jansson R. (2003) Global patterns in endemism explained by past climatic change. *Proceedings. Biological sciences / The Royal Society*, **270**, 583–90.

Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Kawecki T.J. & Holt R.D. (2002) Evolutionary consequences of asymmetric dispersal rates. *The American Naturalist*, **160**, 333–347.

Keenan K., Mcginnity P., Cross T.F., Crozier W.W., & Prodöhl P.A. (2013) DiveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, **4**, 782–788.

Kimmel T.M., do Nascimento L.M., Piechowski D., Sampaio E.V.S.B., Rodal M.J.N., & Gottsberger G. (2010) Pollination and seed dispersal modes of woody species of 12-year-old secondary forest in the Atlantic forest region of Pernambuco, NE Brazil. *Flora*, **205**, 540–547.

Koch I., Kinoshita L.S., & Bittrich V. (2007) Taxonomic novelties in *Rauvolfia* (Apocynaceae, Rauvolfioideae) from Brazil. *Novon: A Journal for Botanical Nomenclature*, **17**, 462–471.

Kopelman N.M., Mayzel J., Jakobsson M., Rosenberg N.A., & Mayrose I. (2015) Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, **15**, 1179–1191.

Ku C., Chung W.C., Chen L.L., & Kuo C.H. (2013) The Complete Plastid Genome Sequence of Madagascar Periwinkle *Catharanthus roseus* (L.) G. Don: Plastid Genome Evolution, Molecular Marker Identification, and Phylogenetic Implications in Asterids. *PLoS ONE*, **8**, 1–11.

Leigh, J.W. & Bryant D. (2015). PopART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, **6**, 1110–1116.

Lima J.S., Telles M.P.C., Chaves L.J., Lima-Ribeiro M.S., & Collevatti R.G. (2017) Demographic stability and high historical connectivity explain the diversity of a savanna tree species in the Quaternary. *Annals of botany*, **119**, 645–657.

Lopes A.V. & Machado I.C. (1999) Pollination and Reproductive Biology of *Rauvolfia grandiflora* (Apocynaceae): Secondary Pollen Presentation, Herkogamy and Self-Incompatibility. *Plant Biology*, **1**, 547–553.

Loveless M.D. & Hamrick J.L. (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, **15**, 65–95.

Markgraf V., McGlone M., & Hope G. (1995) Neogene paleoenvironmental and paleoclimatic change in southern temperate ecosystems — a southern perspective. *Trends in Ecology & Evolution*, **10**, 143–147.

Marriage T.N., Hudman S., Mort M.E., Orive M.E., Shaw R.G., & Kelly J.K. (2009) Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity*, **103**, 310–317.

Médail F. & Diadema K. (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. *Journal of Biogeography*, **36**, 1333–1345.

Moraes E.M., Yotoko K.S.C., Manfrin M.H., Solferini V.N., & Sene F.M. (2009) Phylogeography of the cactophilic species *Drosophila gouveai*: Demographic events and divergence timing in dry vegetation enclaves in eastern Brazil. *Journal of Biogeography*, **36**, 2136–2147.

Novaes R.M.L., Lemos Filho J.P., Ribeiro R.A., & Lovato M.B. (2010) Phylogeography of *Plathymenia reticulata* (Leguminosae) reveals patterns of recent range expansion towards northeastern Brazil and southern Cerrados in Eastern Tropical South America. *Molecular Ecology*, **19**, 985–998.

Oliveira-Filho A.T., Ratter J.A. (2002) Vegetation physiognomies and woody flora of the Cerrado biome. In: The Cerrados of Brazil: Ecology and Natural History of a Neotropical Savanna (eds Oliveira PS, Marquis RJ), pp. 91–120. Columbia University Press, New York.

Paradis E. (2010) Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, **26**, 419–420.

Pennington R.T., Prado D.E., & Pendry C.A. (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, **27**, 261–273.

Perez M.F., Bonatelli I.A.S., Moraes E.M., & Carstens B. C. (2016). Model‐based analysis supports interglacial refugia over long‐dispersal events in the diversification of two South American cactus species. *Heredity*, **116**, 550–557.

Peterson, A.T., Soberón, J., Pearson, R.G., Anderson, R.P., Martínez-Meyer, E., Nakamura, M. & Araújo, M.B. (2011). Ecological niches and geographic distributions (Monographs in population biology; no. 49). Princeton University Press. 328 pp.

Phillips S.J., Anderson R.P., & Schapire R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

Prado C.P.A., Haddad C.F.B., & Zamudio K.R. (2012) Cryptic lineages and Pleistocene population expansion in a Brazilian Cerrado frog. *Molecular Ecology*, **21**, 921–941.

Pritchard J.K., Stephens M. & Donnelly P., (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

QGIS Development Team, (2016). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

Rambaut A. (2014). FigTree: A graphical viewer of phylogenetic trees. [WWW document] URL http://tree.bio.ed.ac.uk/software/ [accessed 15 January 2018].

Rambaut A, Suchard M, Drummond A. (2014). Tracer, version 1.6.0. [WWW document] URL http://tree.bio.ed.ac.uk/software/tracer/ [accessed 15 January 2018].

Ramos A.C.S., Lemos-Filho J.P., Ribeiro R.A., Santos F.R., & Lovato M.B. (2007) Phylogeography of the tree *Hymenaea stigonocarpa* (Fabaceae: Caesalpinioideae) and the influence of Quaternary climate changes in the Brazilian cerrado. *Annals of Botany*, **100**, 1219–1228.

Ramos-Onsins S.E. & Rozas J. (2002) Statistical Properties of New Neutrality Tests Against Population Growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rao A. (1956) A revision of *Rauvolfia* with particular reference to the American species. *Annals of the Missouri Botanical Garden*, **43**, 253–354.

Ratter J., Ribeiro J., & Bridgewater S. (1997) The Brazilian Cerrado Vegetation and Threats to its Biodiversity. *Annals of botany*, **80**, 223–230.

Ribeiro P.C., Lemos-Filho J.P., de Oliveira Buzatti R.S., Lovato M.B., & Heuertz M. (2016) Species-specific phylogeographical patterns and Pleistocene east-west divergence in Annona (Annonaceae) in the Brazilian Cerrado. *Botanical Journal of the Linnean Society*, **181**, 21–36.

Rice W.R. (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223.

Richards C.L., Carstens B.C., & Lacey Knowles L. (2007) Distribution modelling and statistical phylogeography: An integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography*, **34**, 1833–1845.

Ronquist F., Teslenko M., Van Der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A., & Huelsenbeck J.P. (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Rosenberg N.A. (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Sanaiotti T.M., Martinelli L.A., Victoria R.L., Trumbore S.E., & Camargo P.B. (2002) Past vegetation changes in amazon savannas determined using carbon isotopes of soil organic matter. *Biotropica*, **34**, 2–16.

Santos M.G., Nogueira C., Giugliano L.G., & Colli G.R. (2014) Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography*, **41**, 1506–1519.

Schoener T.W. (1968) Anolis lizards of Bimini: resource partitioning in a complex fauna. *Ecology* **49**, 704–726.

Simões A.O., Kinoshita L.S., Koch I., Silva M.J., & Endress M.E. (2016) Systematics and character evolution of Vinceae (Apocynaceae). *Taxon*, **65**, 99–122.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Turchetto-Zolet A.C., Pinheiro F., Salgueiro F., & Palma-Silva C. (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–1213.

Vieira F. de A., Novaes R.M.L., Fajardo C.G., Santos R.M. dos, Almeida H. de S., Carvalho D. de, & Lovato M.B. (2015) Holocene southward expansion in seasonally dry tropical forests in South America: phylogeography of *Ficus bonijesulapensis* (Moraceae). *Botanical Journal of the Linnean Society*, **177**, 189–201.

Warren D.L., Glor R.E., Turelli M. (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868-2883.

Werneck F.P. (2011) The diversification of eastern South American open vegetation biomes: Historical biogeography and perspectives. *Quaternary Science Reviews*, **30**, 1630–1648.

Werneck F.P., Nogueira C., Colli G.R., Sites J.W., & Costa G.C. (2012) Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *Journal of Biogeography*, **39**, 1695–1706.

Yamane K., Yano K., & Kawahara T. (2006) Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. *DNA Research*, **13**, 197–204.

Supplementary data

Table 1: Mutation-scaled effective population size (Θ) and mutation-scaled immigration rates (*M*) for populations of *Rauvolfia weddelliana* and *Rauvolfia gracilis* calculated with Migrate-n.

		Immigration rates						
Population	Θ	$M_{\rightarrow SV}$	$M_{\rightarrow CD}$	$M_{\rightarrow PA}$	$M_{\rightarrow VN}$	$M_{\rightarrow CP}$	$M_{\rightarrow VL}$	$M_{\rightarrow RP}$
SV	0.89	-	6.10	4.39	2.06	2.00	1.53	1.82
		477.4		470.1				662.3
CD	136.80	0	-	2	622.04	487.49	500.41	1
								553.4
PA	137.14	82.40	613.30	-	468.71	515.56	472.42	3
		538.5						
VN	141.76	6	470.39	3.10	-	481.60	401.46	1.82
		207.4		507.8				492.5
CP	122.24	6	553.68	3	534.13	-	3.17	1
		466.6		453.6				
VL	105.12	9	487.55	0	541.95	609.81	-	1.82
RP	0.72	1.82	320.72	3.70	2.27	2.03	2.22	-

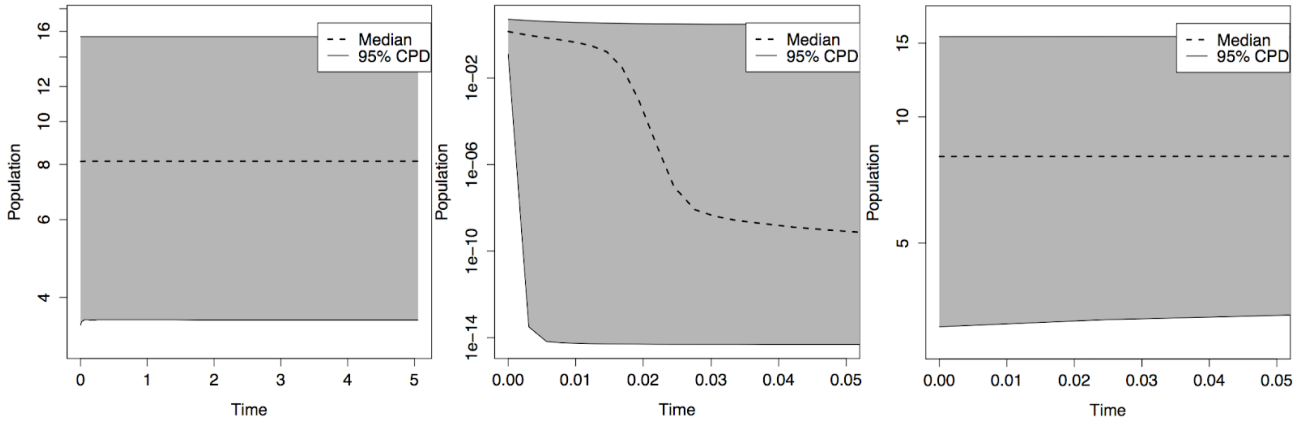


Figure 1: Bayesian Coalescent Skyplot for clades I, II, and III, respectively. Population axis is the log scale population size. Time axis represents million years before present.

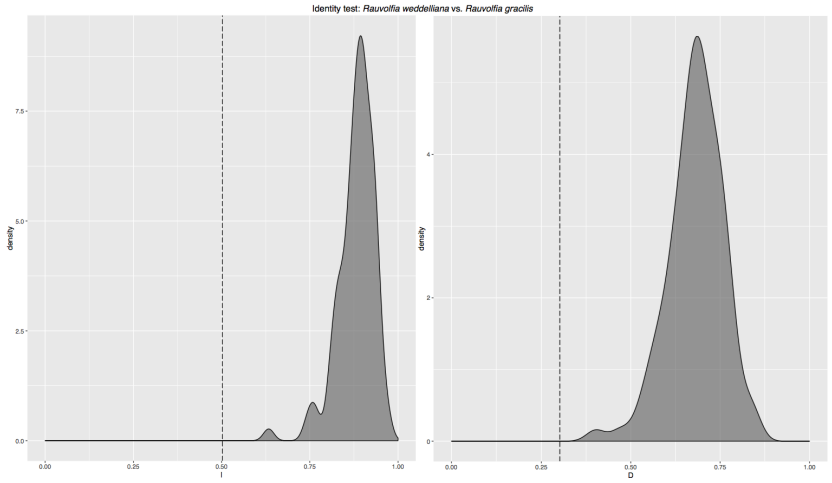


Figure 2: Niche identity tests Warren's I (Warren *et al*. 2008) (left) and Schoener's D (Schoener 1968) (right). Grey curves represent the distribution of neutral model generated by combination of all points, while dashed lines represent actual data statistics.

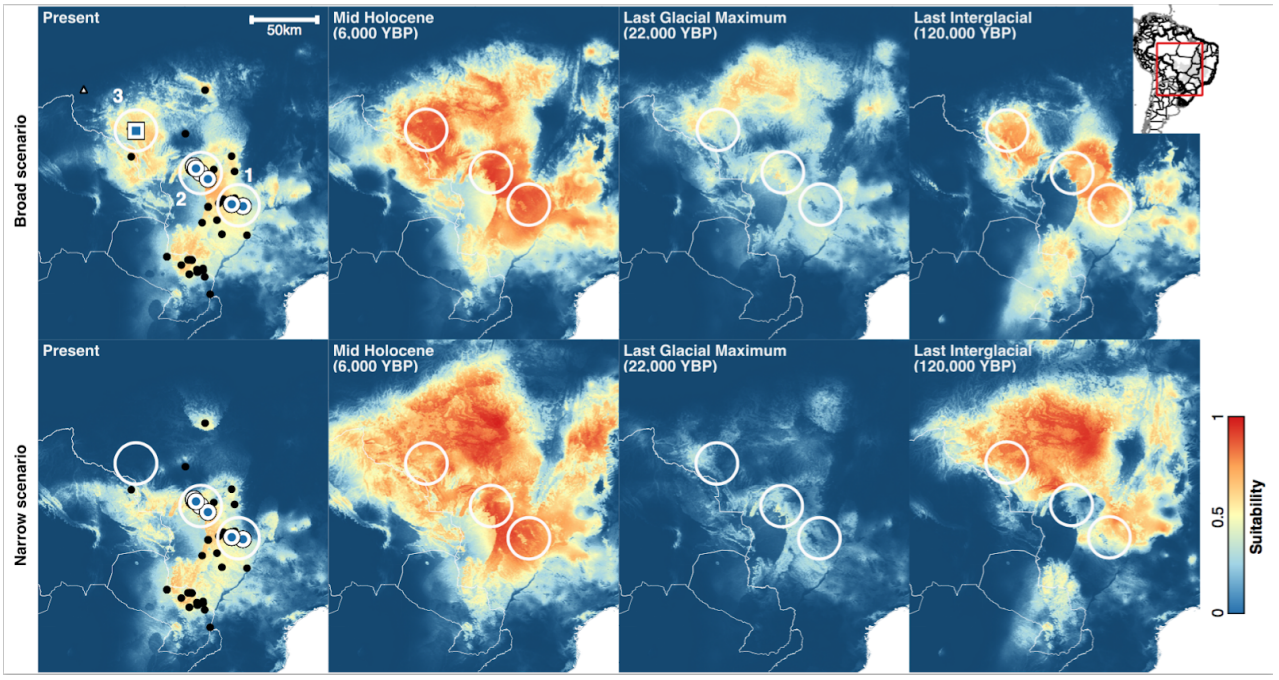


Figure 3: MAXENT models for *Rauvolfia weddelliana* Müll.Arg. and *Rauvolfia gracilis* I. Koch & Kin.-Gouv. potential distribution for present, mid-Holocene (6,000 years before present) and Last Glacial Maximum (21,000 years before present). White and blue colored circles and square indicate sampled localities for *R. gracilis* and *R. weddelliana*, respectively. Black triangles and circles represent database occurrence points for *R. gracilis* and *R. weddelliana*, respectively, used for generating the distribution models. Colder colors indicate low suitability, while hot colors indicate high values of suitability.

4. The impacts of landscape composition, marginality and climate stability on the patterns of endemism of Cerrado woody plants

Vidal, J. D.[1]*, Souza, A. P., Koehler, S.[2] & Koch, I.[2]

[1]Universidade Estadual Paulista "Júlio de Mesquita Filho" – Instituto de Biociências de Botucatu -Departamento de Botânica. Rua Prof. Dr. Antônio Celso Wagner Zanin, s/nº. Rubião Júnior – Botucatu - SP. CEP 18618-689.

[2]Universidade Estadual de Campinas, Departamento de Biologia Vegetal. Rua Monteiro Lobato, 255. Cidade Universitária "Zeferino Vaz" - Barão Geraldo - Campinas – SP. Caixa Postal 6109 CEP 13083-970.

*Email: jdvidal@ibb.unesp.br

Abstract

Many theories were proposed to explain the outstanding endemism of plants in Cerrado. Among the different diversification mechanics suggested, the most widely discussed are the plateau/valley, stable/unstable, and core/peripheral hypothesis. The first argues that plateaus harbor more ancient lineages than peripheral valley regions, and therefore should present higher levels of endemism. The second theory suggests that the stability of the climate in certain regions promoted more adequate environments for the maintenance of paleoendemics species. The last scenario attributes the distribution of endemism levels in Cerrado to gradients of optimal conditions available to locally adapted species, predicting

higher endemism levels in core regions of the domain than in marginal areas. In this study, we compared the endemism patterns of woody plants of Cerrado with the predictions of each theory to discuss the importance of each of these processes in the modern distribution of endemism nuclei along the domain. We generated an endemism map for 174 species of woody plants, comprising most vegetation types of Cerrado. We used spatial analysis tools, species distribution modelling and tests of correlation to summarize the importance of each predictor on endemism. We found correlations for endemism, elevation, and marginality, supporting the plateau/valley and core/peripheral hypothesis. We demonstrated that, even though plateaus are more stable climatically environments; their elevation, not their climatically stability, predict endemism patterns. Our past vegetation reconstructions with simulated datasets supported the expansion of Cerrado plateau vegetation and the retraction of vegetation in valleys during Last Glacial Maximum. However, real occurrence data models also indicated that this climatic stability scenario might be too simplistic, since most plant species in valleys presented contrasting responses to past climatic shifts. Our results contribute to the understanding of processes shaping endemism patterns in Cerrado, and may also assist on the interpretation of the variable response of historical distributions of its species reported by previous studies.

Key-words: Biogeography. Species distribution modelling. Community ecology.

Introduction

Cerrado is the largest savanna and the second major bioregion in South America, with a complex and unique biogeographic history. Not only Cerrado harbors more plant species

than any other savanna in the world (>7000; Mendonça *et al*. 1998) but also, more than 44% of its flora is endemics, making it the richest tropical savanna in the world in this aspect (Klink & Machado 2005). The landscape of Cerrado is divided in ancient plateaus, which are large extension areas with relative high elevation (500–1700 m), and peripheral valleys, which are lower elevation regions (100–500 m) that surround highlands (Silva 1997, Silva & Bates 2002). These landscape unities present different biotic composition and geological histories, indicating that their biogeographical histories were also profoundly distinct. The biota of peripheral valleys is heterogeneous, composed by gallery forests, riverine forests, marshlands, tropical dry forests, and enclaves of species from adjacent biomes, like the Amazon and the Atlantic Rainforests (Silva & Bates, 2002). Plateaus, on the other hand, present a more dominant vegetation type, composed majorly by savannahs, which range from open fields ("*campo limpo*") to dense forest type vegetation ("*cerradão*") (Silva 1997).

The outstanding diversity of plants species found in Cerrado is a topic of crescent discussion in Neotropical biogeography. Many theories were proposed to explain the high degree of endemism found within the Cerrado for several taxa, but due to the variety of responses of organisms and the gap of studies concerning this environment, no clear pattern of the biogeographic history of the domain have emerged so far (Turchetto-Zolet *et al*. 2013). Concerning the modern discussed theories, three models of diversification are among the mostly supported scenarios: (1) the geographic compartmentalization of the landscape in highlands and depressions of the Late Miocene as drivers of allopatric speciation between disjoint landscape units; (2) the differences in Pleistocene climatic stability along regions within the domain promoting the maintenance of diversity in stable areas; and (3) the gradient of optimal environmental conditions between core and peripheral Cerrado allowed higher

population densities in nuclear areas, favoring the maintenance of endemics in these regions. Although they are not mutually excluding, the relative importance of these theories remain a major focus of debate, since different patterns of organisms seem to support different scenarios (e.g. Turchetto-Zolet *et al*. 2013).

The first model is based on the fact that plateaus are old geological formations with final uplift by the Late Tertiary, while valleys are resultant of more recent erosive processes (Ab'Saber 1983, Silva 1997, Werneck 2011). Thus, the ancient geomorphological processes of these environments may have promoted the isolation of older lineages in highlands, while the recently formed valleys were occupied by neoendemics species (Silva, 1997). This model is consistent with endemism patterns reported for birds (Silva, 1997), plants of the genus *Mimosa* (Simon & Proença, 2001), and squamate lizards (Nogueira *et al*. 2011).

Along with the geomorphological processes of late Miocene, the Pleistocene climate was also proposed as a major driver of the diversification of species and lineages within Cerrado (Werneck *et al*. 2012, Santos *et al*. 2014). This theory is closely associated with the plateau/valley hypothesis, since peripheral valleys are often assumed to be more sensible to Quaternary climatic shifts, due to their environmental heterogeneity and the frequent presence of species from adjacent biomes (Ab'Sáber 1983). In this scenario, during cycles of cold and dry climate, the distribution of moist vegetation from valley environments underwent strong retractions. Contrastingly, the vegetation of plateaus expanded into valleys, possibly even penetrating the modern distribution of adjacent domains (van der Hammen, 1974). Indeed, climatic stable regions are demonstrated to predict diversity in both interspecific (Werneck 2011) and intraspecific levels (Carnaval, 2000; Carnaval & Moritz, 2008), as areas presenting less stable environmental conditions are more prone to face local extinctions. Also, as climate

changes, organisms locally adapted respond by migrating or adapting to the new climatic conditions (Aitken *et al*. 2008).

The third model argues that core regions of Cerrado historically presented more optimal conditions for the maintenance of populations, and thus supported the diversification of endemics species through time (Eckert *et al*. 2008). In this scenario, core regions of Cerrado are expected to show higher levels of endemism, while peripheral regions should present sub optimal conditions, and thus, were most likely to have underwent stochastic events of local extinction than core regions. Some examples of organisms supporting this scenario are the intraspecific diversity of squamate lizards of the genus *Micrablepharus* (Santos *et al*. 2014).

Multiple species approaches are useful to detect general patterns, which are important to assemble large-scale scenarios. Large-scale hypothesis testing for identification of refugia has been held for several regions around the world, with important biogeographic outcomes (Thuiller 2004; Svenning *et al*. 2008). Mostly, single species patterns or simulated dataset models have been applied in past distributional reconstructions for Cerrado (Werneck *et al*. 2012), while multiple species approaches less common (Siqueira & Peterson 2003; Bonaccorso *et al*. 2006, Bueno *et al*. 2017). Also, no simulations explicitly incorporated the differences in abiotic and biotic composition of valleys and plateaus so far. Examples of community level studies using multiple species models are found for other Neotropical biomes, like SDTFs (Werneck *et al*. 2011), Atlantic Rainforest (Colombo & Joly 2010), and Amazon Rainforest (Bonaccorso *et al*. 2004). Multiple species reconstructions approach for Cerrado are desirable, since they would allow investigations about the response its different vegetation types to past and future climatic shifts. Siqueira & Peterson (2003) applied

community-level models to test the responses of 163 species of trees of Cerrado to future climatic changes, but in a straight conservation biology approach. Similar studies focused on historical patterns would allow important insights on the biogeographic history of the domain, as performed by Bueno *et al*. (2017).

In this study, we compared the levels of endemism of the woody flora of Cerrado with the three scenarios proposed to explain its diversification. We generated maps of distribution of plants species occurring in savanna, gallery forest, riverine forest and seasonally dry tropical forest within Cerrado. Then, we used this information to generate maps of endemism for the domain. We applied linear models to test the correlation of the endemism with altitude, stability and marginality data. Also, to test specifically the convergence of climatic stability and the plateau/valley hypothesis, we compared variance of environmental conditions of valleys and plateaus during the past 120,000 years. We used real species occurrence data of woody plants from valley and plateaus to test the scenarios of retraction/expansion of different vegetation types occurring in these landscape units during glacial cycles, in a community level approach. Our results contribute to the discussions about the scenarios of Quaternary climatic shifts in Cerrado vegetation, which must necessarily consider the importance of the heterogeneity of responses of the different vegetation types of the domain and the variable climatic history and composition of its landscape units.

Methods

Species occurrence data and endemism calculation

To generate the maps of endemism for woody species of Cerrado, we gathered occurrence data for 63 savanna species, 54 seasonally dry tropical forest species, 36 gallery

species, and 21 riverine forest species (Table 1). We based our selection on a consistent classification system adopted for defining vegetation types (Oliveira-Filho, 2017). We elaborated species list for savanna, gallery and riverine forests based on NeoTropTree database (Oliveira-Filho, 2017). For SDTFs, we adopted the catalogue proposed by Prado and Gibbs (1993). Savanna list included species considered present in Cerrado arid woodlands, while gallery forest species list included solely species present in Cerrado riverine forests and riverine forest. Riverine forests and gallery forests are considered different vegetation types due to their composition and the restricted distribution of riverine forests in association with major rivers of the domain (Ribeiro & Walter 1998; Silva & Bates 2002). We excluded species present in multiple vegetation types in order to reduce the bias of widespread species in models. We retrieved species occurrence data from GBIF using the R package "Rgbif" (Chamberlain *et al*. 2016). To ensure the quality of data, we adopted search parameters that restrict data without original coordinates and with spatial issues (parameters: *hasCoordinate* = *TRUE*; *hasGeospatialIssue* = *FALSE*). We manually checked points by plotting individual maps and discarding outliers. We also discarded species with poor occurrence information (<10 registries). We generated a map of weighted endemism using the algorithm proposed by Guerin et al. (2015). We used a geographic grid with 0.25 degrees cells to generate a presence/absence matrix. Then, we calculated the number of unique species that occur in each cell, and weighted this statistic for each species, by dividing it by the number of grid cells in which they occur (Guerin *et al*. 2015). By doing so, we were able to reduce the bias of widespread species in our endemism estimates.

Endemism correlation with elevation, marginality and climate

To evaluate the relative climatic stability of plateaus and valleys, we divided the whole Cerrado domain in highlands (> or = 500m of elevation) and depressions (< 500m of elevation). We randomly sampled 1000 points for each landscape unity (Figure 1a and 1b) and extracted values of 19 WorldClim bioclimatic variables (Hijmans *et al*. 2005) to compare the present environmental conditions between areas. To compare climatic stability, we sampled the same random points for past climatic scenarios (mid Holocene, Last Glacial Maximum and Last Interglacial - Hijmans *et al*. 2005). We calculated variance for environmental variables of plateaus and valleys and compared variances using a bootstrap approach, by sampling 1000 repetitions of the variances and comparing the significance of outcomes with Welch's t-test (Welch 1938) for each variable in valley and plateaus.

To generate layers of climatic stability, we performed a principal component analysis (PCA) with the variances obtained for the bioclimatic layers. Then, we transformed the most informative components (cumulative proportion >75%) in spatial layers with R package "raster" (Hijmans & Etten 2014).

To test the correlation of endemism with elevation and marginality, we extracted values of elevation and distance from the Cerrado shape centroid for 1000 random points within Cerrado. Then, we extracted values of endemism for these same points and fitted linear models to check for correlation between these parameters. Finally, we applied a Pearson's product-moment correlation test to estimate the significance of these associations.

Historical vegetation distribution

To test the historical climate hypothesis, we tested if plateaus vegetation expanded or retracted during glaciations, and if valley response was opposite, as proposed for the refugia

scenario. We generated distribution models for simulated points of valleys and plateaus and also for real occurrence of species associated with gallery forests, riverine forests, seasonally dry tropical forests and savanna (Table 1). To select climatic variables used in niche models, we randomly sampled 1000 points from all 19 Worldclim bioclimatic variables (Hijmans 2001) and calculated Pearson's pairwise correlation. Variable pairs with high correlation (> 0.75 or <-0.75) were randomly selected and discarded. We selected a subset of 12 less correlated variables: annual precipitation, precipitation seasonality, precipitation of driest quarter, precipitation of warmest quarter, precipitation of coldest quarter, mean diurnal range, isothermality, temperature seasonality, max temperature of warmest month, mean temperature of wettest quarter, mean temperature of driest quarter, and elevation. For past scenarios (mid Holocene and Last Glacial Maximum), we used the same bioclimatic variables from MIROC-ESM and CCSM4 scaled climatic models (Hijmans *et al*. 2005). We averaged models between both scenarios (MIROC-SM and CCSM4) and presented consensus for each time period. For last Interglacial (LIG; ~120,000 - 140,000 years BP), we also adopted bioclimatic values from WorldClim, which was based on the reconstruction proposed by Otto-Bliesner *et al*. (2008).

We used Maxent v 3.3.3 (Phillips *et al*. 2006) to generate species distribution models. We process models and bioclimatic layers in R with packages "dismo" (Hijmans *et al*. 2012) and "raster" (Hijmans & Etten 2014). For species distribution models, we used 1000 background points for each species, in order to estimate background information with pseudo-absences. Models performance was evaluated by AUC-ROC values. We converted present models into binary models by establishing a threshold based on the lower presence training (LPT) value from which a presence point is recorded on the actual data for the present

model (Pearson *et al*. 2007). Then, we applied this same procedure to past scenarios, using the present model threshold. We summarized information from models by counting per species presence cells for each scenario reconstruction. To facilitate comparison between species, we converted the values into percentages by dividing areas by the larger range registered for each species. With this, we were able to detect in which climatic scenario a species reached the maximum range of its distribution and to compare current and LGM distributions of species. We also combined all binary models, both for each species and for the whole dataset, in order to identify individual and community level stable areas. For the simulated datasets, the same approach was implemented.

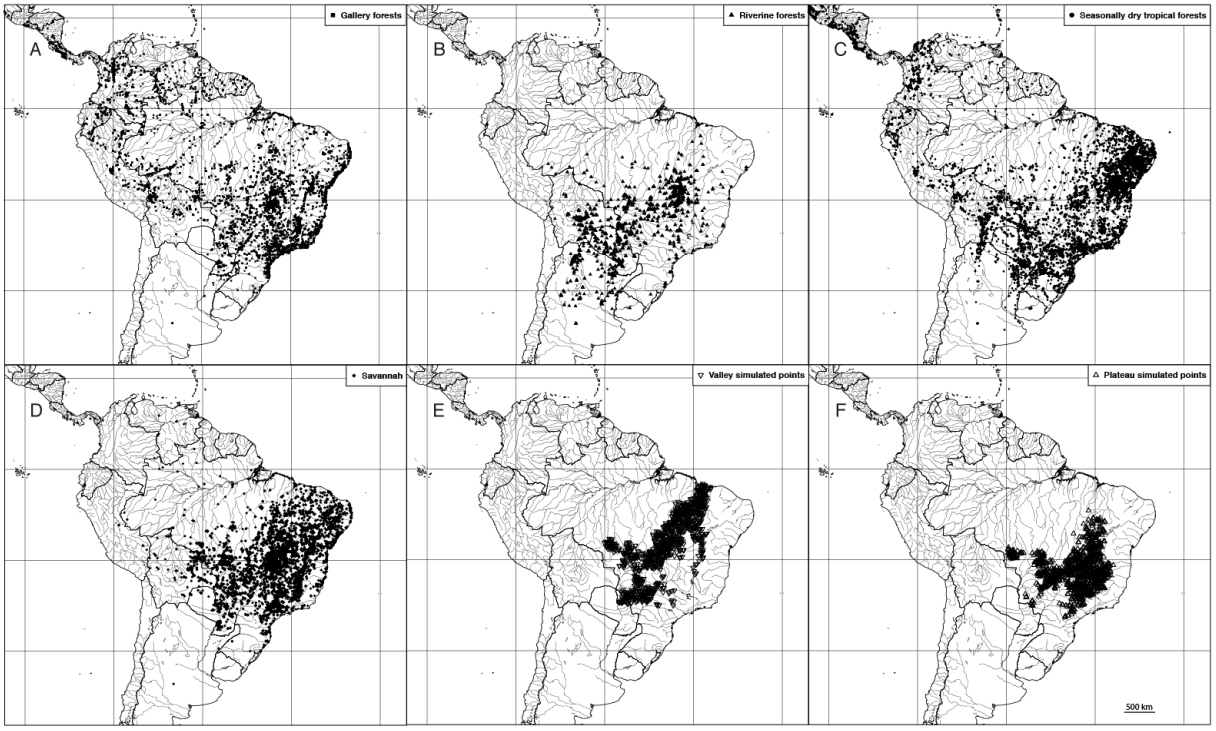


Figure 1: Points used for spatial analysis of Cerrado and species distribution models. (A) Simulated points for depressions (< 500m elevation); (B) Simulated points for plateaus and highlands (> or = 500m elevation); (C) Occurrence points for 21 species of gallery forests used in distribution models; (D) Occurrence points for 63 species of savanna used in distribution models; (E) Simulated points for depressions (< 500 m elevation); (F) Simulated points for plateaus and highlands (> or = 500 m elevation).

Results

Species list and endemism calculation

In our survey, we obtained 9452 registries for 54 seasonally dry tropical forest species; 1085 occurrence records for 21 riverive forest species; 6903 records for 36 gallery forest species. For savanna species (Cerrado sensu stricto), we recovered and 9145 occurrences for 63 savanna species. Species list and number of occurrences for each one are listed in Supplementary data, Appendix 1, 2, 3 and 4. Our estimates of weighted endemism for the Cerrado domain returned higher levels of endemism in the states of Minas Gerais, Goiás and Bahia (Figure 2). By comparing the endemism map we produced with the official map of landscape unities for Brazil, we were able to identify endemism nuclei in (1) Chapada dos Guimarães, (2) Vale do Araguaia, (3) Serra do Roncador, (4) Serra dos Pirineus, (5) Chapada dos Veadeiros, (6) Planalto do Distrito Federal, (7) Chapada do Rio São Francisco and (8) Serra do Cipó.

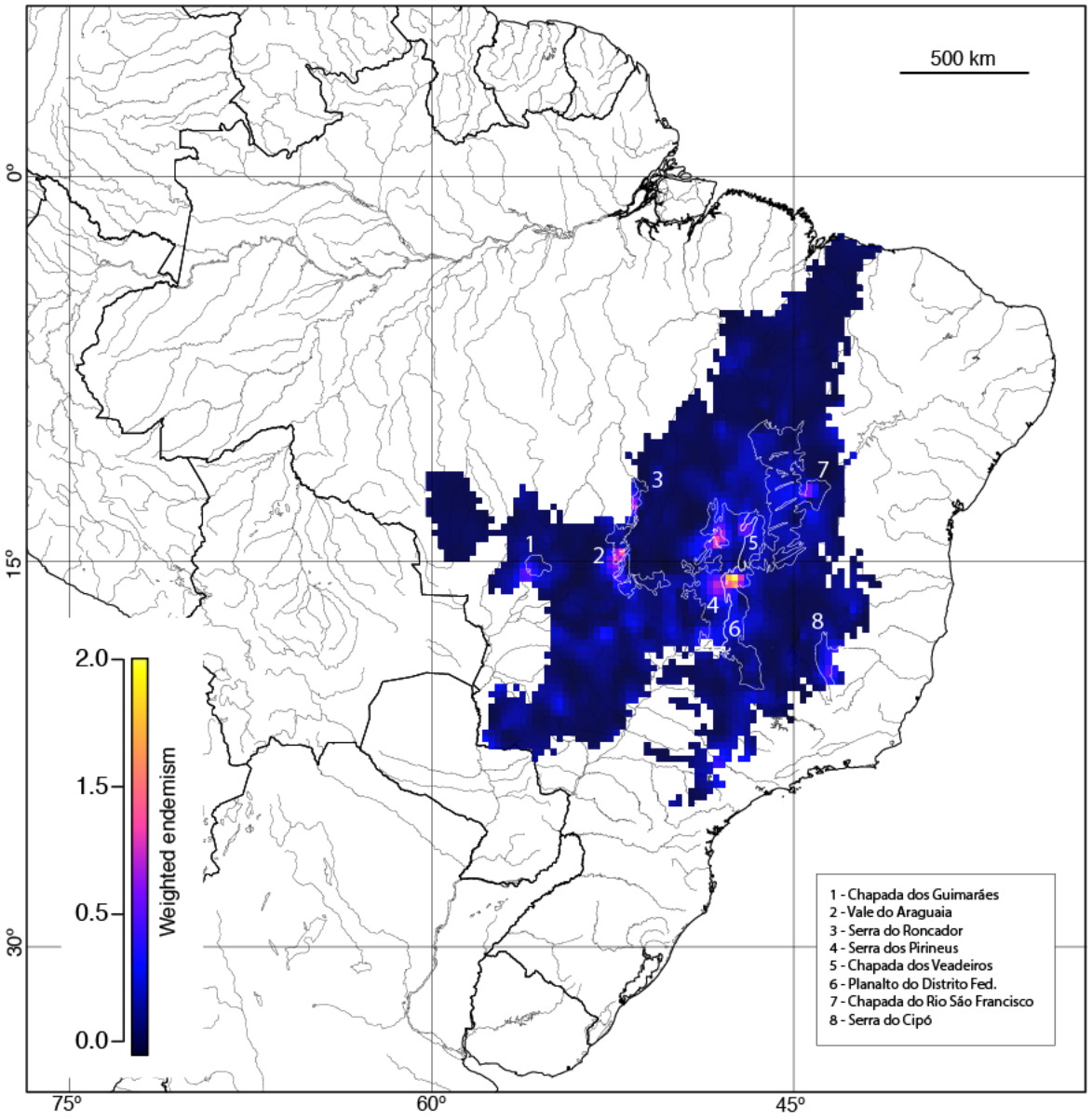


Figure 2: Weighted endemism map for woody plants in the Cerrado, calculated with Global Biodiversity Information (GBIF) data. Color-scale is proportional to the value of weighted endemism (Guerín *et al*. 2015) per cell.

Endemism correlation with elevation, marginality and climate

All 19 bioclimatic variables analyzed presented significant differences between valleys and plateaus (Figure 3). A total of 16 variables (84%) presented lower variances for plateaus than for valley, indicating that plateau climatic conditions were more stable along the time

periods that we compared. Other three variables displayed lower variances in valleys. Welch's two sample t-test returned p-values lower than 0.01 for all comparisons.

Principal component analysis of the variances of all bioclimatic layers indicated five most informative components (Table 1), which summarize most variation observed for variances of environmental layers in four climatic scenarios analyzed. All variance layers and PCA raster layers are available in Supplementary data (Appendix 5 and 6).

Table 1: Principal Component Analysis loadings of bioclimatic layers variance for the most informative components (>0.75) based on cumulative proportion. Red tones indicate strength of positive correlations, and blue tones indicate strength of negative correlations.

Environmental variable	**Comp.1**	**Comp.2**	**Comp.3**	**Comp.4**	**Comp.5**
Annual Mean Temperature	-0.411	-0.170	-0.060	-0.111	0.104
Mean Diurnal Range	-0.237	0.203	-0.181	0.048	0.528
Isothermality	-0.169	-0.376	0.090	-0.059	-0.296
Temperature Seasonality	0.131	-0.018	-0.374	-0.278	0.299
Max Temperature of Warmest Month	0.267	-0.148	-0.386	0.029	0.006
Min Temperature of Coldest Month	0.010	-0.281	0.280	-0.129	0.328
Temperature Annual Range	0.322	-0.217	-0.210	0.043	0.039
Mean Temperature of Wettest Quarter	0.253	-0.162	-0.397	-0.022	-0.043
Mean Temperature of Driest Quarter	-0.043	-0.336	0.261	-0.142	0.262
Mean Temperature of Warmest Quarter	-0.248	0.164	-0.146	-0.018	-0.302
Mean Temperature of Coldest Quarter	-0.213	0.293	-0.149	0.007	0.008
Annual Precipitation	0.068	0.068	-0.008	-0.629	-0.171
Precipitation of Wettest Month	0.204	0.234	0.192	-0.239	0.264
Precipitation of Driest Month	0.144	-0.186	0.323	0.123	0.171
Precipitation Seasonality	-0.123	0.367	0.018	0.113	0.170
Precipitation of Wettest Quarter	-0.318	-0.180	-0.097	-0.355	-0.148
Precipitation of Driest Quarter	0.129	0.256	0.148	-0.500	-0.014
Precipitation of Warmest Month	-0.157	-0.205	-0.312	-0.083	0.273
Precipitation of Coldest Month	-0.395	-0.142	-0.068	0.014	0.097

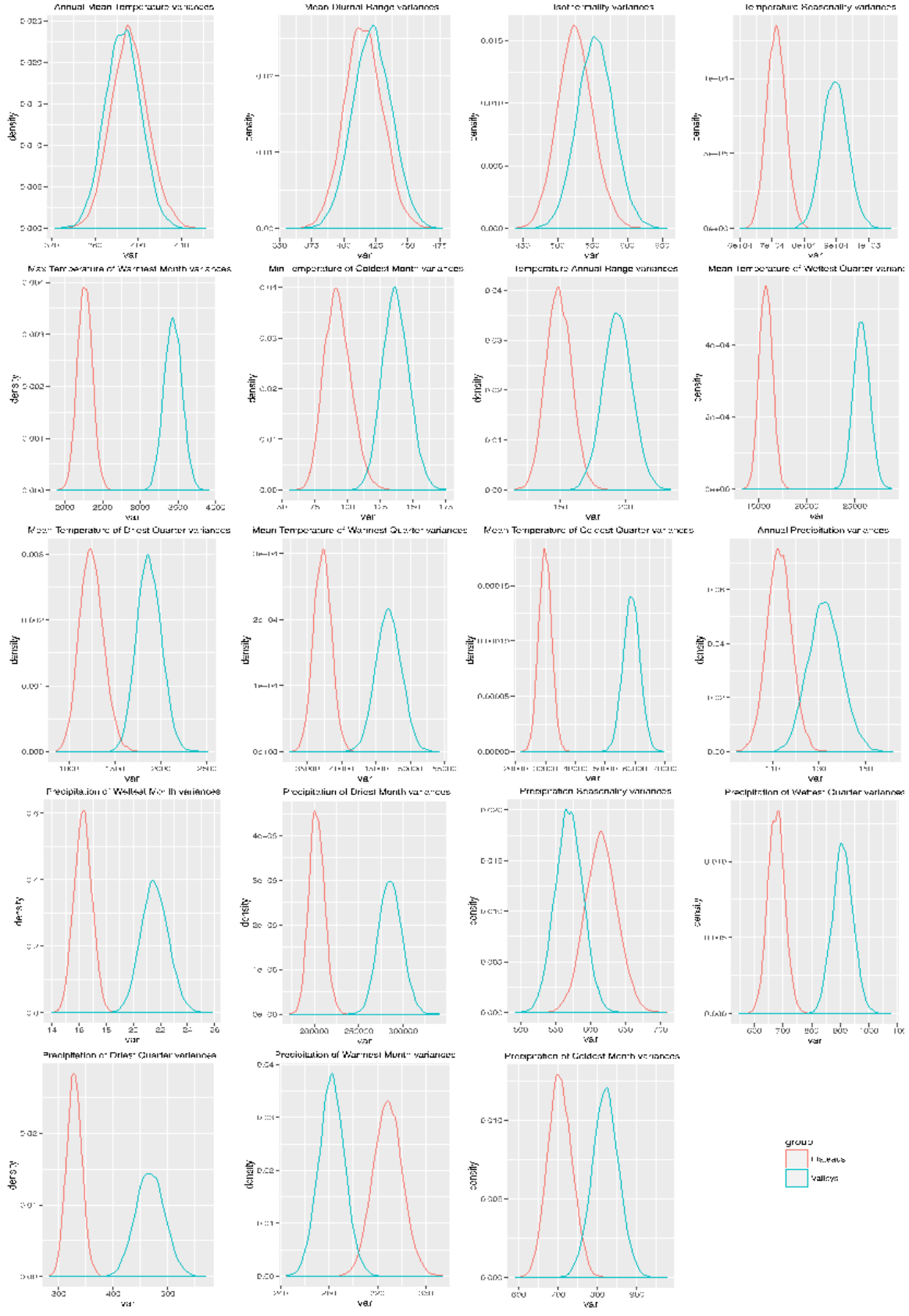

Annual Mean Temperature variances
Mean Diurnal Range variances
Isothermality variances
Temperature Seasonality variances
Max Temperature of Warmest Month variances
Min Temperature of Coldest Month variances
Temperature Annual Range variances
Mean Temperature of Wettest Quarter variances
Mean Temperature of Driest Quarter variances
Mean Temperature of Warmest Quarter variances
Mean Temperature of Coldest Quarter variances
Annual Precipitation variances
Precipitation of Wettest Month variances
Precipitation of Driest Month variances
Precipitation Seasonality variances
Precipitation of Wettest Quarter variances
Precipitation of Driest Quarter variances
Precipitation of Warmest Month variances
Precipitation of Coldest Month variances
density
var
group
Uplands
Valleys

Figure 3: Bootstrap distribution of variances during last 120,000 years for 19 bioclimatic variables of Cerrado valleys and plateaus.

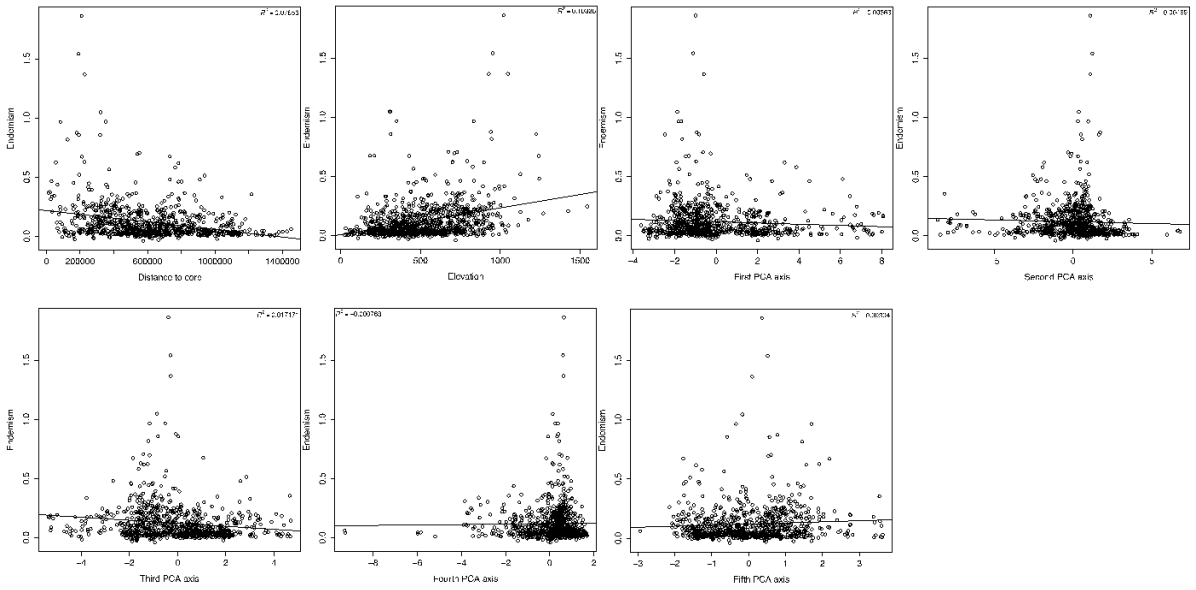


Figure 4: Correlation and linear regression between weighted endemism, marginality, elevation, and stability axis. R^2 values represent the coefficient of determination. Correlation was significant for endemism vs. distance to core, endemism vs. elevation and endemism vs. third PCA axis (p-value < 0.01).

Positive correlation was observed for elevation and endemism, while distance to core and endemism presented negative correlation (Figure 4). We found significant correlation between weighted endemism and marginality, elevation and the third PCA axis of the stability variances (Table 2). We identified no correlation between endemism and the other four PCA axes. The Pearson's product-moment correlation tests recovered significant p-values for marginality, elevation and third axis of the stability PCA (Table 2).

Table 2: Correlation between weighted endemism and each predictor tested. R^2: values of coefficient of determination. (*) *p*-values > 0.01, tested with Pearson's product-moment correlation.

Predictor	R^2	***p*-value**
Marginality	0.077	*
Elevation	0.103	*
Stability		
PCA1	0.004	
PCA2	0.005	
PCA3	0.017	*
PCA4	0.001	
PCA5	0.003	

Historical vegetation distribution

AUC for simulated datasets models was 0.736 for plateaus and 0.722 for valleys. For real occurrence data, average AUC values were 0.892 for seasonally dry tropical forest models, 0.956 for riverine forest models, 0.879 for gallery forest models and 0.915 for savanna models, indicating high predictive accuracy of models. Our simulated models recovered a clear pattern of glacial expansion for plateaus, while valleys presented glacial

retraction in suitability (Figures 3 and 4). In fact, the lowest proportional distribution for valleys and the highest for plateaus took place during Last Glacial Maximum (Figure 4).

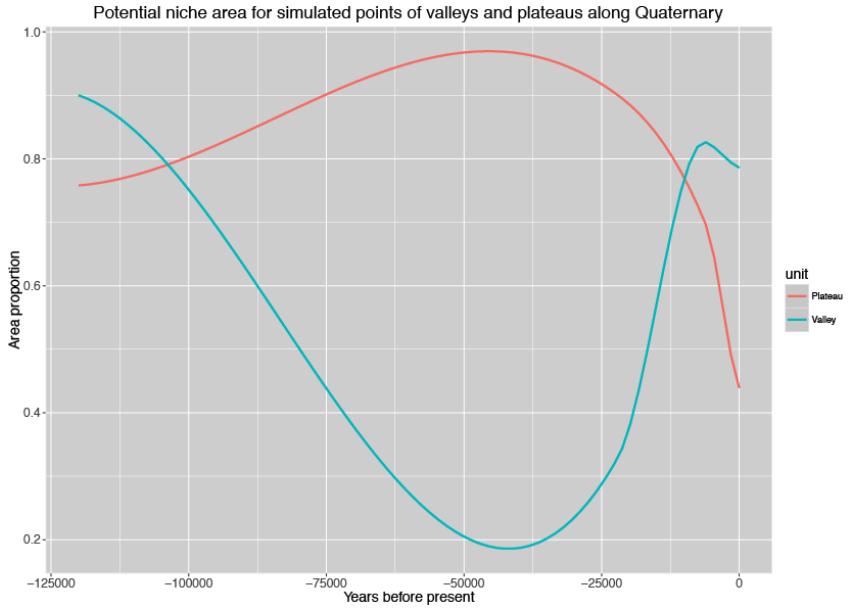


Figure 3: Proportional area of potential niches for plateau and valley environments along the last interglacial and glacial maximum. Area was estimated by converting distribution models to binary models based on the lowest suitability found for points of the present model, divided by the widest modeled scenario range

We found that 38% of savanna species presented their maximum range during glacial periods, while gallery species, riverine species and SDTF species presented variable patterns (Supplementary data, Appendix 1, 2, 3 and 4). For the forest vegetation types, 81% of the modeled riverine species displayed wider distribution ranges during interglacial periods. SDTFs and gallery forests presented an opposite pattern to the riverine forests, with 72% of gallery forest species and 54% of SDTF species presenting their wider distribution during LGM. For savanna, we found that 65.08% of species presented wider LGM ranges than present distribution. Combined models for all species binary distributions resulted in community level reconstructions of vegetation types (Figure 4). For these combined

reconstructions, we also observed a similar response of wider occurrence of SDTFs and gallery forests during LIG and LGM, while savannas presented a more local expansion within Cerrado. Riverine forests, in comparison, presented a slightly narrower distribution during LGM, and were greatly restricted during LIG.

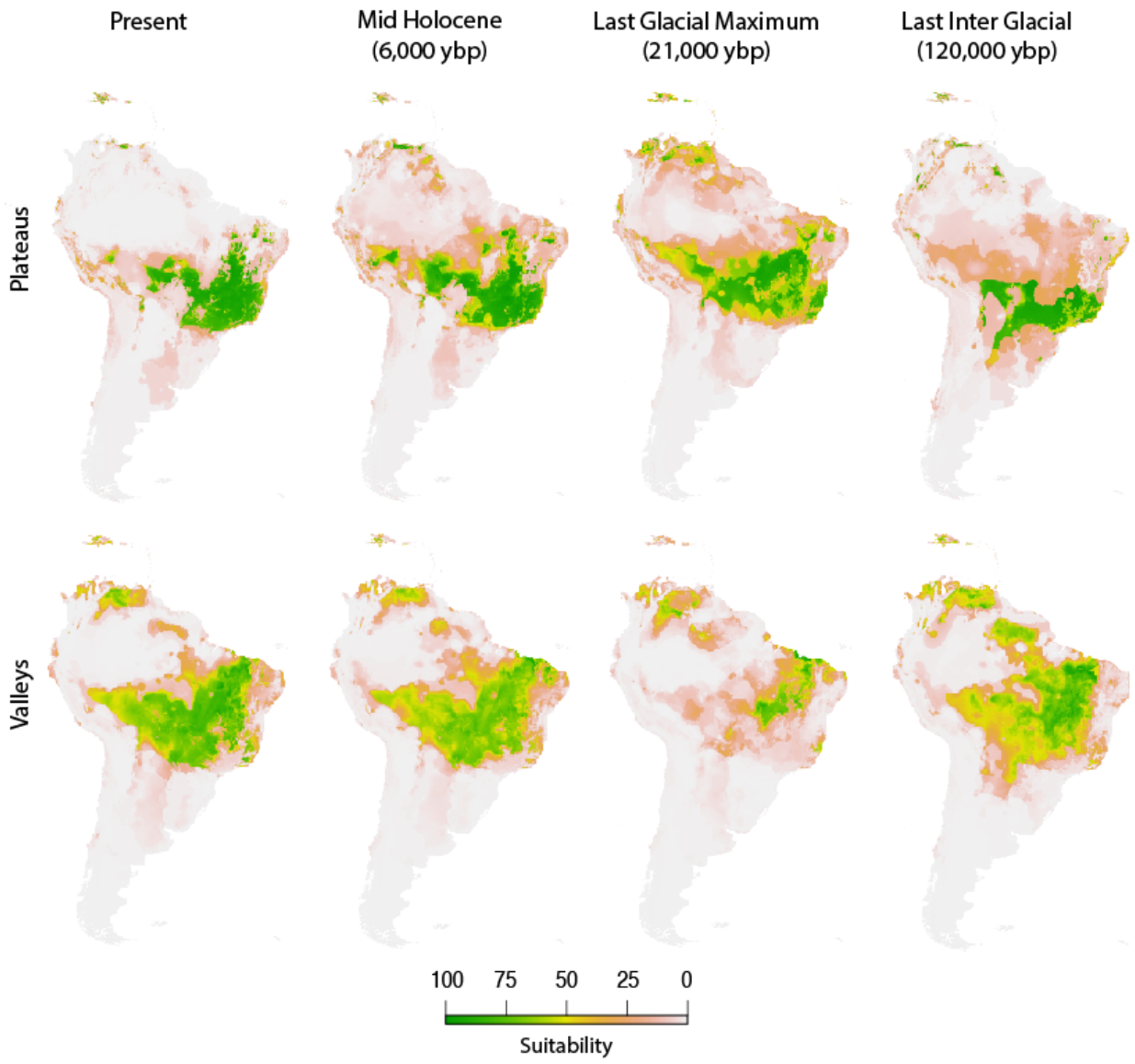


Figure 4: Distribution models based on simulated datasets (1000 random points) for Cerrado plateaus (elevation > or = 500 m) and valleys (elevation < 500 m) for present, mid Holocene (6,000 years before present), Last Glacial Maximum (21,000 years before present) and Last Inter Glacial (120,000 years before present).

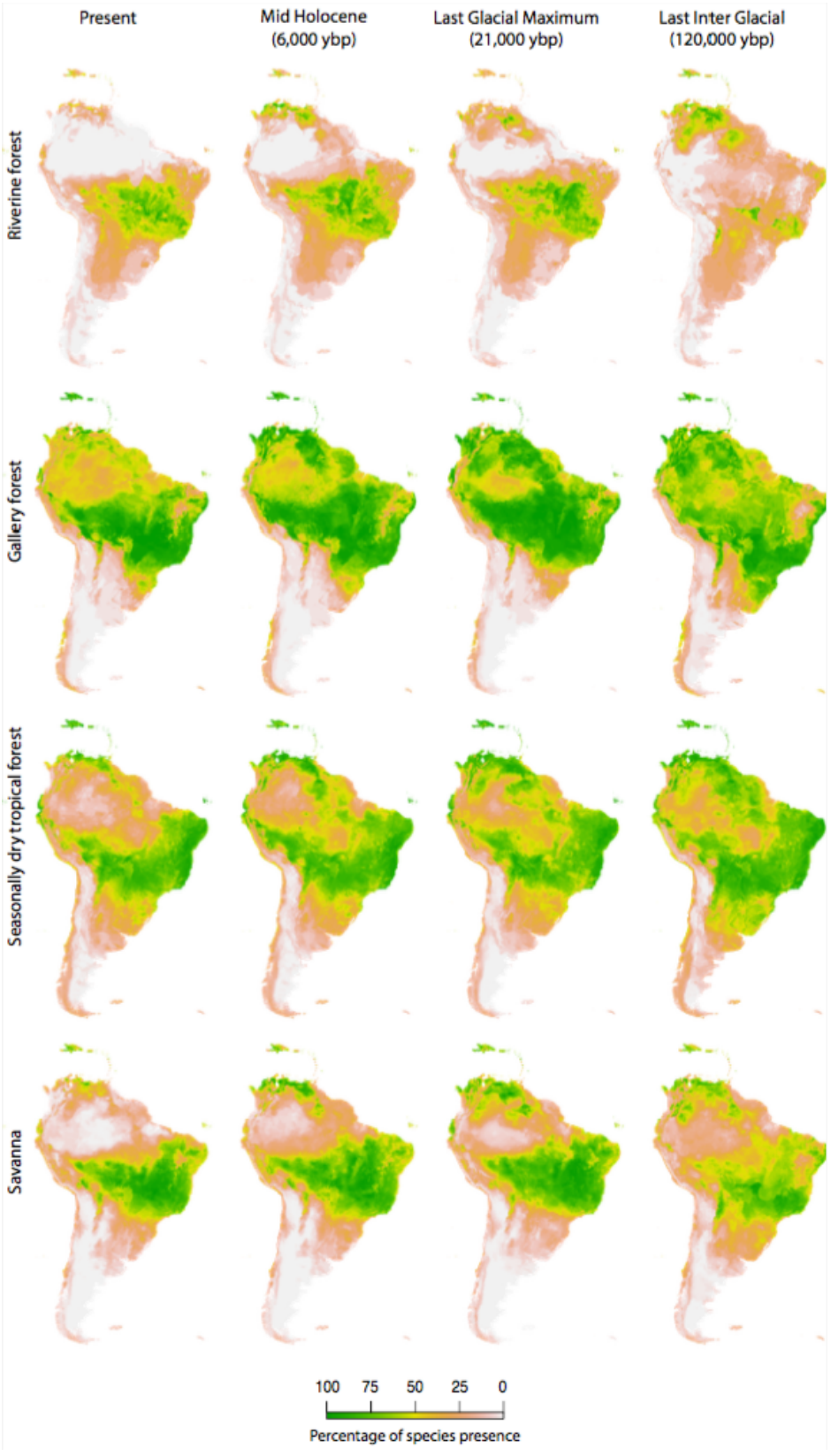

Present
Mid Holocene
(6,000 ybp)
Last Glacial Maximum
(21,000 ybp)
Last Inter Glacial
(120,000 ybp)
Riverine forest
Gallery forest
Seasonally dry tropical forest
Savanna
100 75 50 25 0
Percentage of species presence

Figure 5: Community level models for South America riverine forests, gallery forests, seasonally dry tropical forests, and savanna for present, mid Holocene (6,000 years before present), Last Glacial Maximum (21,000 years before present) and Last Inter Glacial (120,000 years before present).

Discussion

In this study, we identified regions concentrating the endemism of plants species for the Cerrado domain and tested its relationship with elevation, marginality and stability. We found significant correlation of endemism with elevation and marginality. Also, we demonstrated that climate in plateau environments presented an overall higher stability than valley environments. This stability, however, showed low correlation with endemism. We also generated vegetation distribution models for Cerrado, which showed different patterns of responses for simulated datasets and real occurrence data. While our simulated dataset models fitted the previously reported scenario of glacial expansion in arid Cerrado (Werneck 2012), with plateaus expanding and valleys retracting on Last Glacial Maximum, real occurrence data models showed a variety of responses. Even though most savanna species presented expansions during glaciation, the same pattern was recovered for gallery and seasonally dry tropical forests. We only recovered glacial retraction for riverine forests valley vegetation. This suggests that the plateau/valley compartmentalization may be too simplistic for the heterogeneity of vegetation composition within the Cerrado domain.

The identification of nuclei of endemism in Chapada dos Guimarães, Chapada dos Veadeiros, Serra do Cipó, and Distrito Federal is consistent with patterns reported by other researches (Simon & Proença 2000, Nogueira *et al*. 2011). Most of these areas are high elevation, with a single valley region (Vale do Araguaia) recovered as highly endemic. This patterns of endemism associated with highlands is consistent with the results showed by our extant analysis of correlations. Elevation and landscape compartmentalization are considered

important drivers of diversification, since differences in elevation might promote isolation, which is a key process for speciation. Although the concentration of endemism levels for plants in high elevations have been already suggested in a global scale (Steinbauer *et al*. 2016), a finer scale pattern for Cerrado, like we presented here, reinforces the role of this relationship for this specific domain, as previously suggested (e.g. Fiaschi & Pirani 2009).

Individual species models indicated a contrary range response to the retraction prediction of the "plateaus as refugia" theory (Ab'Saber 1983). Overall, gallery, seasonally dry tropical forests and savanna vegetation expanded during glaciations. Among the vegetation types of peripheral valleys, only riverine forests displayed retractions during glaciations, followed by a mid Holocene expansion. Our models indicated that the retractions occurred during the Last Interglacial may have promoted the habitat fragmentation for savanna, SDTF and gallery forests. Also, our models for SDTF presented a scenario consistent with the glacial expansion scenario proposed by Prado & Gibbs (1993). Previous distribution models for SDTFs, however, refuted the proposal of a wider distribution for this vegetation during glacial periods (Werneck *et al*. 2011). Our results differ from Werneck *et al*. (2011) by suggesting the expansion of SDTFs during LGM, inclusive into eastern corridor of Amazon Basin, a scenario earlier proposed by Pennington *et al*. (2000). For savannas, no clear expansion within other biomes was observed, being the glacial expansion mostly into valley regions and on Tepuis (table mountains) environments, where patches of savanna are still found nowadays (Fiaschi & Pirani 2009). We observed very similar patterns for simulated and real occurrence datasets of savannas and plateau points, confirming the close association of this vegetation with highlands in Cerrado. Our results disagree with the pattern of widespread savanna scenario in LIG with posterior retraction proposed by Bueno *et al*. (2017). Their

models recovered a retraction on suitability of plateaus during the LGM, which was interpreted as an evidence of valleys role as refugia. For our models, the decrease The gradual retraction observed from glaciations to present is consistent with the scenario of interglacial refugia, commonly proposed for highland restricted taxa in Cerrado, like *Pilosocereus aurisetus* (Bonatelli *et al*. 2014) and *Tibouchina papyrus* (Collevatti *et al*. 2012). Our results reinforce the argument that effects of interglacial periods in the distributions of Cerrado species might be considered a key processes for their diversification.

The plateaus comprised areas with higher climatic stability for most of environmental variables included in our analysis. It is important to notice, however, that valleys also presented lower variances for some components, like annual mean temperature, precipitation seasonality, and precipitation of the warmest month, indicating that these conditions are more stable for these environments. Taxa associated with plateaus or valleys are not restricted to its environments, since distribution of species is dependent on the ecological plasticity of organisms. Thus, even highly specialized species are expected to occur in different habitats (e.g. Hortal *et al*. 2009). We demonstrate here that even though typical vegetation of valleys is composed by riverine, gallery and seasonally dry forests, species of these vegetation-types are not restricted to these environments and their historical distribution responses are independent.

Along with elevation, we recovered marginality as an important predictor of endemism. This is explained by the fact that species that are adapted to a certain environment tend present higher survivability, reproductive success and population growth in the optimal conditions (Eckert *et al*. 2008). These conditions are gradually distributed along space, and are expected to be more available in nuclear areas of distribution of organisms. This result is

also supported by the higher historical suitability of nuclear region habitats to plants of Cerrado proposed by Bueno *et al*. (2010). However, studies refuting the assumption of optimal performances for nuclear populations indicated that this pattern might not be strongly supported for plants (Abeli *et al*. 2014). Therefore, it is still a topic of controversy the real importance of interactions between marginality and endemism. A caveat in our results, however, is related to the inclusion criterion we adopted. Since we restricted our dataset to include only species with a reasonable number of samples to apply in the modelling step, we may have favored the inclusion of species of wider distributions within the domain, for which the optimal conditions might be coincident with nuclear regions of Cerrado. So, we are conservative about this result and recommend further studies, including intraspecific traits, like population sizes and genetic diversity measures, to test the extent of the pattern we recovered. Overall, our results provided a general view of the relative impacts of these theories in a complex environment like Cerrado. We recommend that further studies with other groups of organisms, in community-level approaches, as we implemented here, may allow the comparison of the extension of our results into the unique and threatened biota of Cerrado.

Conclusion

We provided endemism maps for Cerrado woody plants, and compared the pattern of endemism with biogeographical theories of Cerrado, which assumed plateau/depression, stability/instability and core/peripheral processes on the generation of endemism. Also, we reconstructed the distribution of these species and environments for the recent Quaternary climate. Our estimates of endemism distribution are more likely explained by elevation and

marginality, providing an important insight on the biogeographic processes that generated the Cerrado diversity. We concluded that savannas expanded during last glacial, while valley vegetation presented variable responses. Also, we found evidences that plateaus are more stable scenarios, for most climatic variables. We found weak evidence, however, that this stability may be associated with endemism.

References

Ab'Sáber A.N. (1983) O domínio dos cerrados: introdução ao conhecimento. *Revista do Serviço Público*, **111**, 41–55.

Abeli T., Gentili R., Mondoni A., Orsenigo S., & Rossi G. (2014) Effects of marginality on plant population performance. *Journal of Biogeography*, **41**, 239–249.

Aitken S.N., Yeaman S., Holliday J.A., Wang T., & Curtis-McLane S. (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Bonaccorso E., Koch I., & Peterson A.T. (2006) Pleistocene fragmentation of Amazon species' ranges. *Diversity and Distributions*, **12**, 157–164.

Bonatelli I.A.S., Perez M.F., Peterson A.T., Taylor N.P., Zappi D.C., Machado M.C., Koch I., Pires A.H.C., & Moraes E.M. (2014) Interglacial microrefugia and diversification of a cactus species complex: Phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. *Molecular Ecology*, **23**, 3044–3063.

Bueno M.L., Pennington R.T., Dexter K.G., Kamino L.H.Y., Pontara V., Neves D.M., Ratter J.A., & de Oliveira-Filho A.T. (2017) Effects of Quaternary climatic fluctuations on the distribution of Neotropical savanna tree species. *Ecography*, **40**, 403–414.

Carnaval A., Hickerson M., Haddad C., Rodrigues M., & Moritz C. (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, **323**, 785–789.

Carnaval A.C. & Moritz C. (2008) Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, **35**, 1187–1201.

Chamberlain S., Ram K. Barve V., and Mcglinn D. (2016) rgbif: Interface to the Global Biodiversity Information Facility API. R package version 0.9.2.

Collevatti R.G., de Castro T.G., Lima J. de S., & Telles M.P. de C. (2012) Phylogeography of *Tibouchina papyrus* (Pohl) toledo (Melastomataceae), an endangered tree species from rocky savannas, suggests bidirectional expansion due to climate cooling in the Pleistocene. *Ecology and Evolution*, **2**, 1024–1035.

Colombo A.F. & Joly C.A. (2010) Brazilian Atlantic Forest lato sensu: the most ancient Brazilian forest, and a biodiversity hotspot, is highly threatened by climate change. *Revista Brasileira de Biologia*, **70**, 697–708.

Eckert C.G., Samis K.E., & Lougheed S.C. (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular ecology*, **17**, 1170–88.

Fiaschi P. & Pirani J.R. (2009) Review of plant biogeographic studies in Brazil. *Journal of Systematics and Evolution*, **47**, 477–496.

Guerin G.R., Ruokolainen L., & Lowe A.J. (2015) A georeferenced implementation of weighted endemism. *Methods in Ecology and Evolution*, **6**, 845–852.

Hijmans R.J., Cameron S.E., Parra J.L., Jones P.G., & Jarvis A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hijmans R.J. & van Etten J. (2014). raster: Geographic data analysis and modeling. *R package version*, *2*(8).

Hortal, J., Triantis, K.A., Meiri, S., Thébault, E. & Sfenthourakis, S. (2009) Is- land species richness increases with habitat diversity. *The American Naturalist*, **174**, 205–217.

Klink C.A. & Machado R.B. (2005) Conservation of the Brazilian Cerrado. *Conservation Biology*, **19**, 707–713.

Mendonça R.C., Felfili J.M., Walter B.M.T., Júnior M.C.S., Rezende A.V., Filgueiras T.S., Nogueira P.E. (1998). Flora vascular do Cerrado, 289–556 in Sano S.M., Almeida S.P., (eds). Cerrado: Ambiente e Flora. Brasília: EMBRAPA.

Nogueira C., Ribeiro S., Costa G.C., & Colli G.R. (2011) Vicariance and endemism in a Neotropical savanna hotspot: Distribution patterns of Cerrado squamate reptiles. *Journal of Biogeography*, **38**, 1907–1922.

Oliveira-Filho, A. T. 2017. NeoTropTree, Flora arbórea da Região Neotropical: um banco de dados envolvendo biogeografia, diversidade e conservação.

Otto-Bliesner B.L. (2006) Simulating arctic climate warmth and icefield retreat in the Last Interglaciation. *Science*, **311**, 1751–1753.

Pearson R.G., Raxworthy C.J., Nakamura M., & Townsend Peterson A. (2007) Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34**, 102–117.

Pennington R.T., Prado D.E., & Pendry C.A. (2000) Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, **27**, 261–273.

Prado, D.E. & Gibbs, P.E. (1993) Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden*, **80**, 902–927.

Santos M.G., Nogueira C., Giugliano L.G., & Colli G.R. (2014) Landscape evolution and phylogeography of *Micrablepharus atticolus* (Squamata, Gymnophthalmidae), an endemic lizard of the Brazilian Cerrado. *Journal of Biogeography*, **41**, 1506–1519.

Silva J.M.C. da (1997) Endemic bird species and conservation in the Cerrado Region, South America. *Biodiversity and Conservation*, **6**, 435–450.

Silva J.M.C. da & Bates J.M. (2002) Biogeographic Patterns and Conservation in the South American Cerrado: A Tropical Savanna Hotspot. *BioScience*, **52**, 225.

Simon M.. & Proença C. (2000) Phytogeographic patterns of Mimosa (Mimosoideae, Leguminosae) in the Cerrado biome of Brazil: an indicator genus of high-altitude centers of endemism? *Biological Conservation*, **96**, 279–296.

Siqueira M.F. De & Peterson A.T. (2003) Consequences of global climate change for geographic distributions of Cerrado tree species. *Biota Neotropica*, **3**, 1–14.

Svenning J.C., Normand S., & Kageyama M. (2008) Glacial refugia of temperate trees in Europe: Insights from species distribution modelling. *Journal of Ecology*, **96**, 1117–1127.

Thuiller W. (2004) Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, **10**, 2020–2027.

Turchetto-Zolet A.C., Pinheiro F., Salgueiro F., & Palma-Silva C. (2013) Phylogeographical patterns shed light on evolutionary process in South America. *Molecular Ecology*, **22**, 1193–1213.

van der Hammen T. (1974) The Pleistocene Changes of Vegetation and Climate in Tropical South America. *Journal of Biogeography*, **1**, 3.

Welch B.L. (1938) The Significance of the Difference Between Two Means when the Population Variances are Unequal. *Biometrika*, **29**, 350.

Werneck F.P., Costa G.C., Colli G.R., Prado D.E., & Sites Jr J.W. (2011) Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on

palaeodistribution modelling and palynological evidences. *Global Ecology and Biogeography*, **20**, 272–288.

Werneck F.P., Nogueira C., Colli G.R., Sites J.W., & Costa G.C. (2012) Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. *Journal of Biogeography*, **39**, 1695–1706.

Supplementary data

Appendix 1: List of modeled species for seasonally dry tropical forests in South America and their proportional distribution range during last 120,000 years. List of species for each formation was based on Prado & Gibbs (1993). *N*: number of occurrences; *AUC*: Area under the curve; *Present*: proportional range on present; *MH*: proportional range on mid Holocene (6,000 ybp); *LGM*: proportional range on Last Glacial Maximum (22,000 ybp); *LIG*: proportional distribution during Last Interglacial (130,000 ybp); $R_{LGM}>R_P$: range of LGM distribution wider than modern range.

			Proportional distribution				
Seasonally dry forest species	N	AUC	Present	MH	LGM	LIG	$R_{LGM}>R_P$
Albizia inundata	21	0.981	1	0.72	0.076	0.943	
Alseis floribunda	228	0.889	0.727	0.912	0.98	1	Yes
Amburana cearensis	259	0.848	0.718	0.893	0.8	1	Yes
Anadenanthera colubrina	347	0.838	0.722	0.9	0.763	1	Yes
Aspidosperma cuspa	129	0.899	0.999	0.976	0.694	1	
Aspidosperma discolor	121	0.903	0.728	1	0.991	0.971	Yes
Aspidosperma polyneuron	141	0.900	0.828	0.812	1	0.905	Yes
Aspidosperma pyrifolium	326	0.855	0.796	0.901	0.643	1	
Aspidosperma riedelii	24	0.965	0.805	0.87	1	0.948	Yes
Astronium fraxinifolium	295	0.830	0.705	0.873	0.766	1	Yes
Balfourodendron riedelianum	166	0.908	0.628	0.744	1	0.726	Yes
Brunfelsia australis	78	0.938	0.774	0.773	1	0.582	Yes
Brunfelsia uniflora	230	0.864	0.72	0.909	0.882	1	Yes
Calycophyllum multiflorum	139	0.911	1	0.927	0.565	0.957	
Carica quercifolia	24	0.942	0.566	0.606	1	0.631	Yes
Celtis pubescens	171	0.884	0.782	0.85	0.881	1	Yes
Combretum leprosum	324	0.843	0.793	0.829	0.67	1	
Cordia alliodora	157	0.866	1	0.931	0.909	0.919	
Cordia glazioviana	12	0.968	0.57	0.41	0.536	1	
Cordia trichotoma	327	0.830	0.933	0.977	0.957	1	Yes
Couepia uiti	80	0.925	0.688	0.965	0.232	1	
Coutarea hexandra	271	0.844	0.798	0.917	0.876	1	Yes
Diatenopteryx sorbifolia	126	0.922	0.745	0.816	1	0.804	Yes
Diplokeleba floribunda	79	0.944	0.602	1	0.477	0.866	
Enterolobium contortisiliqum	326	0.817	0.958	0.974	0.953	1	
Fraunhofera multiflora	50	0.968	0.711	0.603	0.572	1	
Hymenaea courbaril	258	0.824	0.955	0.976	0.912	1	
Hymenaea eriogyne	89	0.941	0.632	0.496	0.625	1	
Ipomoea carnea	150	0.871	0.92	0.952	0.737	1	
Machaerium acutifolium	302	0.833	0.654	0.882	0.886	1	Yes

Maytenus truncata	52	0.939	0.766	1	0.702	0.86	
Myracrodruon urundeuva	55	0.908	0.651	0.854	0.759	1	Yes
Myroxylon balsamum	154	0.864	0.989	0.969	0.982	1	
Parkinsonia aculeata	34	0.967	0.661	0.623	0.395	1	
Peltophorum dubium	319	0.837	0.765	0.919	0.865	1	Yes
Phyllostylon rhamnoides	78	0.931	0.948	1	0.438	0.792	
Phytolacca dioica	191	0.885	0.586	0.665	1	0.769	Yes
Piptadenia viridiflora	243	0.870	0.873	0.949	0.849	1	
Platypodium elegans	318	0.830	0.81	0.89	0.909	1	Yes
Poeppigia procera	170	0.881	0.894	0.984	1	0.973	Yes
Pouteria gardneriana	109	0.901	0.774	1	0.911	0.944	Yes
Pterogyne nitens	315	0.850	0.587	0.924	0.628	1	Yes
Rauvolfia ligustrina	104	0.913	0.798	0.904	0.503	1	
Ruprechtia laxiflora	270	0.854	0.857	1	0.998	0.967	Yes
Sarcomphalus cotinifolius	62	0.947	0.583	0.788	0.489	1	
Sarcomphalus joazeiro	14	0.967	1	0.585	0.089	0.699	
Schinopsis brasiliensis	307	0.864	0.578	0.987	0.845	1	Yes
Senna spectabilis	297	0.848	0.642	0.801	1	0.922	Yes
Sideroxylon obtusifolium	254	0.869	0.873	0.968	0.819	1	
Solanum granuloso-leprosum	252	0.871	0.75	0.916	0.913	1	Yes
Spondias tuberosa	302	0.870	0.539	0.883	0.608	1	Yes
Sterculia striata	197	0.864	0.877	0.924	0.782	1	
Tabebuia aurea	26	0.944	0.734	1	0.494	0.854	
Varronia leucocephala	79	0.941	0.511	0.541	0.589	1	Yes

Appendix 2: List of modeled species for riverine forests in South America and their proportional distribution range during last 120,000 years. List of species for each formation was based on NeoTropTree database (Oliveira-Filho, 2017). *N*: number of occurrences; *AUC*: Area under the curve; *Present*: proportional range on present; *MH*: proportional range on mid Holocene (6,000 ybp); *LGM*: proportional range on Last Glacial Maximum (22,000 ybp); *LIG*: proportional distribution during Last Interglacial (130,000 ybp); $R_{LGM}>R_P$: range of LGM distribution wider than modern range.

			Proportional distribution				
Riverine forest species	N	AUC	Present	MH	LGM	LIG	$R_{LGM}>R_P$
Acosmium cardenasii	42	0.958	**1**	0.946	0.789	0.659	
Annona aurantiaca	47	0.966	**1**	0.676	0.15	0.487	
Aspidosperma quebracho-blanco	157	0.904	0.913	**1**	0.885	0.978	
Bauhinia mollis	131	0.902	0.794	0.993	0.85	**1**	
Callisthene mollissima	29	0.973	0.876	**1**	0.845	0.51	
Copaifera oblongifolia	78	0.936	0.798	**1**	0.932	0.768	
Eugenia matogrossensis	10	0.926	0.834	0.393	**1**	0.439	Yes

Eugenia ternatifolia	38	0.957	0.377	**1**	0.49	0.93	
Ladenbergia cujabensis	12	0.971	**1**	0.761	0.596	0.474	
Luetzelburgia praecox	19	0.969	0.622	**1**	0.186	0.599	
Lycium cuneatum	38	0.957	0.915	0.874	0.712	**1**	
Machaerium eriocarpum	30	0.972	0.692	0.755	0.08	**1**	
Mimosa glutinosa	24	0.987	0.006	0.003	**1**	0.002	Yes
Myrcia camapuanensis	28	0.968	0.836	**1**	0.601	0.437	
Myrcia racemulosa	27	0.975	0.451	**1**	0.686	0.656	
Myrcia tomentosa	25	0.985	0.513	**1**	0.846	0.588	
Pseudobombax minimum	18	0.977	**1**	0.812	0.859	0.411	
Pterodon emarginatus	114	0.924	**1**	0.907	0.956	0.552	Yes
Senegalia praecox	159	0.917	0.882	**1**	0.794	0.675	
Talisia subalbens	14	0.971	0.752	0.781	**1**	0.697	Yes
Vochysia pruinosa	45	0.971	0.85	0.287	**1**	0.251	Yes

Appendix 3: List of modeled species for gallery forests in South America and their proportional distribution range during last 120,000 years. List of species for each formation was based on NeoTropTree database (Oliveira-Filho, 2017). *N*: number of occurrences; *AUC*: Area under the curve; *Present*: proportional range on present; *MH*: proportional range on mid Holocene (6,000 ybp); *LGM*: proportional range on Last Glacial Maximum (22,000 ybp); *LIG*: proportional distribution during Last Interglacial (130,000 ybp); $R_{LGM}>R_P$: range of LGM distribution wider than modern range.

			Proportional distribution				
Gallery forest species	N	AUC	Present	MH	LGM	LIG	$R_{LGM}>R_P$
Apeiba tibourbou	249	0.826	0.953	0.984	0.885	**1**	
Apuleia leiocarpa	284	0.826	0.91	0.957	0.892	**1**	
Callisthene major	152	0.915	0.852	0.975	**1**	0.953	Yes
Calophyllum brasiliense	237	0.821	0.926	0.981	0.909	**1**	
Cardiopetalum calophyllum	164	0.899	0.881	0.898	**1**	0.949	Yes
Cariniana rubra	83	0.935	**1**	0.943	0.716	0.853	
Cheiloclinium cognatum	238	0.832	0.951	0.94	0.929	**1**	
Copaifera langsdorfii	303	0.835	0.677	0.95	0.891	**1**	Yes
Croton urucurana	309	0.824	0.921	0.97	0.967	**1**	Yes
Dendropanax cuneatus	287	0.838	0.681	0.748	**1**	0.677	Yes
Erythroxylum daphnites	204	0.870	0.683	0.854	0.951	**1**	Yes
Euplassa inaequalis	97	0.927	0.935	**1**	0.981	0.915	Yes
Ferdinandusa speciosa	125	0.925	0.788	0.949	0.95	**1**	Yes
Guatteria sellowiana	153	0.920	0.557	0.8	**1**	0.782	Yes
Guettarda viburnoides	309	0.836	0.594	0.864	0.749	**1**	Yes
Hedyosmum brasiliense	261	0.876	0.407	0.69	**1**	0.641	Yes
Licania apetala	277	0.977	**1**	0.928	0.766	0.903	

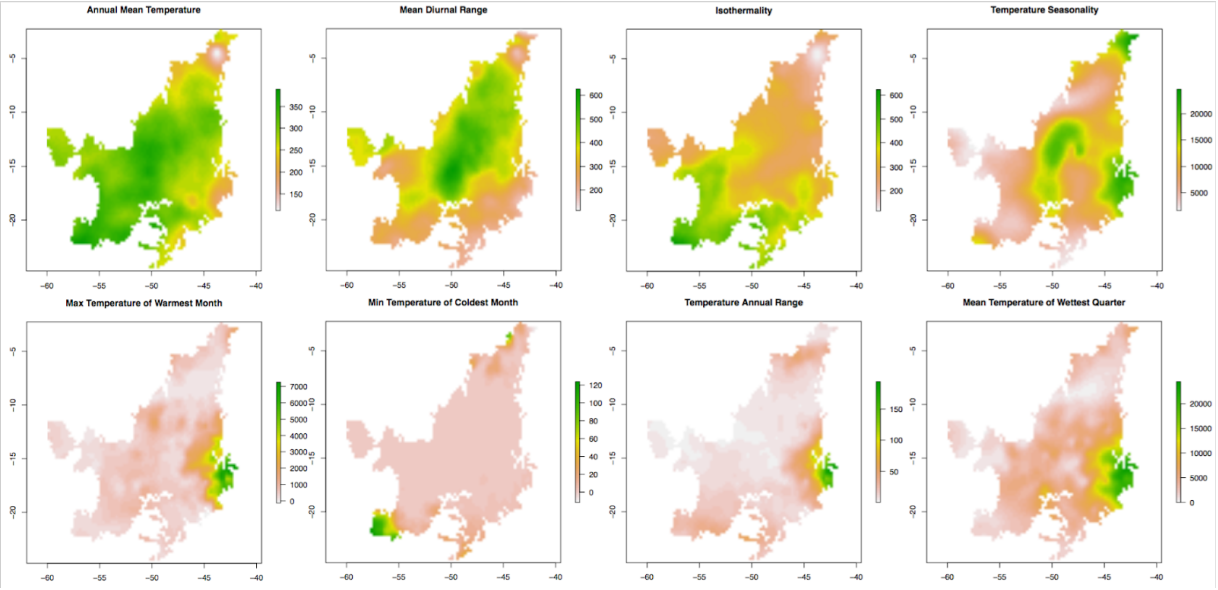
Magnolia ovata	54	0.956	0.464	0.672	**1**	0.457	Yes
Mauritia flexuosa	60	0.841	**1**	0.877	0.769	0.7	
Miconia chartacea	141	0.926	0.558	0.804	**1**	0.707	Yes
Myrcia fenzliana	62	0.947	0.566	0.849	**1**	0.705	Yes
Ocotea aciphylla	201	0.867	0.916	0.979	0.971	**1**	Yes
Piptocarpha macropoda	151	0.923	0.441	0.66	**1**	0.567	Yes
Protium heptaphyllum	240	0.834	0.904	0.942	0.896	**1**	
Pseudolmedia laevigata	266	0.837	**1**	0.991	0.973	0.939	
Richeria grandis	276	0.830	0.952	0.959	0.974	**1**	Yes
Styrax camporum	261	0.866	0.587	0.846	**1**	0.76	Yes
Symplocos nitens	135	0.911	0.515	0.866	0.828	**1**	Yes
Tapirira guianensis	201	0.845	0.923	0.943	0.924	**1**	Yes
Tetragastris balsamifera	43	0.926	0.642	0.909	0.679	**1**	Yes
Virola sebifera	202	0.827	**1**	0.816	0.961	0.795	
Virola urbaniana	18	0.984	0.792	0.801	**1**	0.114	Yes
Vochysia pyramidalis	208	0.887	0.828	0.941	0.894	**1**	Yes
Vochysia tucanorum	273	0.865	0.628	0.831	**1**	0.9	Yes
Xylopia emarginata	136	0.866	0.966	0.965	**1**	0.923	Yes
Xylopia sericea	243	0.836	0.902	0.94	0.916	**1**	Yes

Appendix 4: List of modeled species for savanna in South America and their proportional distribution range during last 120,000 years. List of species for each formation was based on NeoTropTree database (Oliveira-Filho, 2017). *N*: number of occurrences; *AUC*: Area under the curve; *Present*: proportional range on present; *MH*: proportional range on mid Holocene (6,000 ybp); *LGM*: proportional range on Last Glacial Maximum (22,000 ybp); *LIG*: proportional distribution during Last Interglacial (130,000 ybp); $R_{LGM}>R_P$: range of LGM distribution wider than modern range.

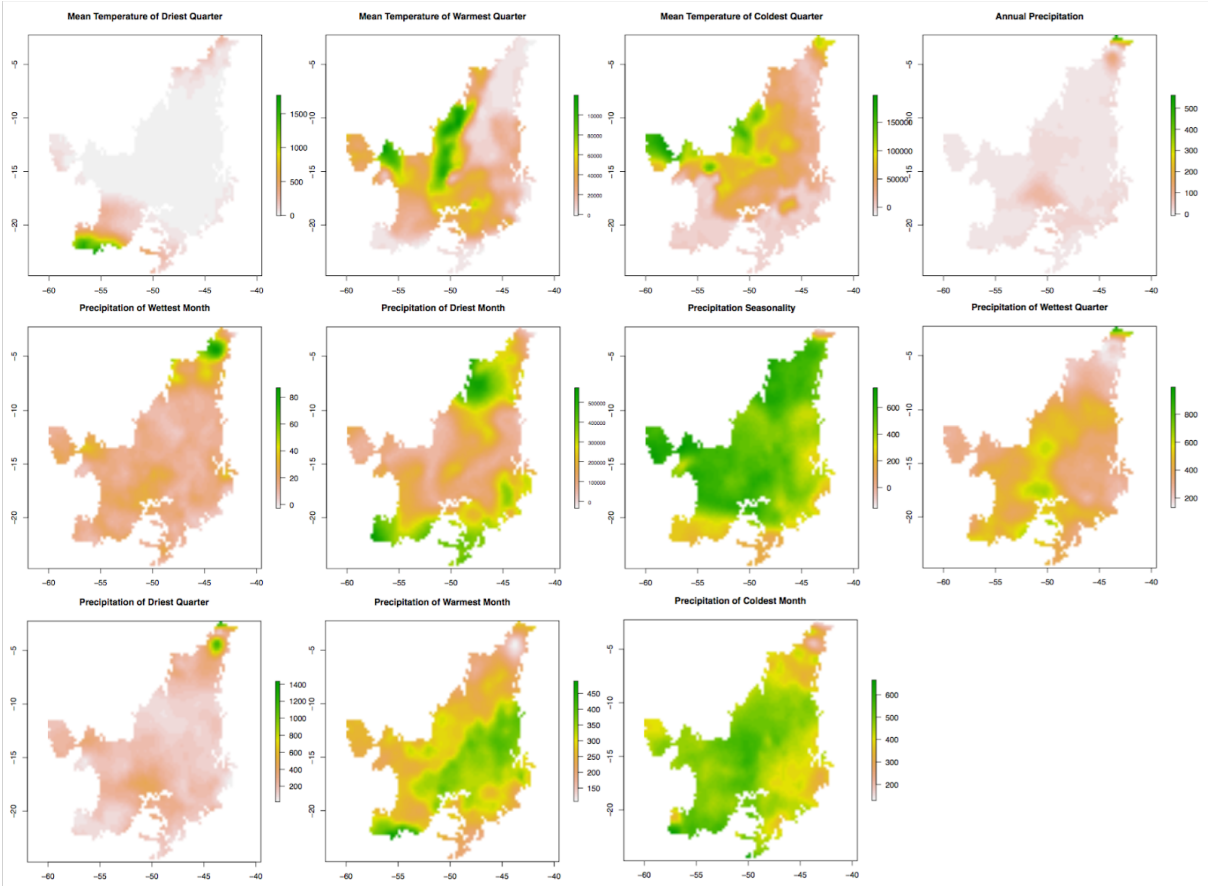
			Proportional distribution				
Savanna species	N	AUC	Present	MH	LGM	LIG	$R_{LGM}>R_P$
Acosmium dasycarpum	315	0.947	0.883	0.938	0.969	**1**	Yes
Aegiphila verticillata	251	0.869	0.666	0.877	**1**	0.987	Yes
Andira cordata	64	0.948	0.707	0.427	0.579	**1**	
Annona coriacea	351	0.822	0.749	0.885	0.881	**1**	Yes
Annona crassiflora	194	0.879	0.74	0.887	0.91	**1**	Yes
Annona nutans	128	0.927	0.936	**1**	0.249	0.512	
Aspidosperma nobile	64	0.918	0.928	**1**	0.746	0.886	
Aspidosperma tomentosum	297	0.848	0.822	0.988	**1**	0.989	Yes
Bauhinia rufa	305	0.845	0.643	0.913	0.882	**1**	Yes
Bursera leptophloeos	317	0.853	0.767	0.898	0.659	**1**	
Caryocar brasiliense	75	0.902	0.778	0.862	**1**	0.766	Yes

Chamaecrista machaeriifolia	33	0.979	0.458	0.932	**1**	0.359	Yes
Chamaecrista orbiculata	163	0.917	0.957	**1**	0.925	0.569	Yes
Connarus detersus	22	0.936	0.676	0.861	0.874	**1**	
Connarus suberosus	280	0.855	0.852	0.915	0.93	**1**	Yes
Copaifera elliptica	33	0.954	0.7	**1**	0.432	0.886	
Copaifera malmei	31	0.964	**1**	0.557	0.233	0.425	
Cordia insignis	86	0.933	0.477	0.817	0.553	**1**	
Dalbergia miscolobium	275	0.872	0.746	0.952	0.957	**1**	Yes
Davilla grandiflora	124	0.902	0.728	0.955	0.936	**1**	Yes
Dimorphandra gardneriana	201	0.886	0.832	0.842	0.866	**1**	
Dimorphandra mollis	292	0.841	0.812	0.908	0.917	**1**	Yes
Diptychandra aurantiaca	171	0.891	0.68	0.865	0.741	**1**	
Eremanthus mattogrossensis	48	0.953	0.754	0.931	**1**	0.776	Yes
Erythroxylum engleri	61	0.935	0.78	0.666	**1**	0.243	Yes
Erythroxylum tortuosum	145	0.920	0.751	0.882	**1**	0.757	Yes
Hancornia speciosa	307	0.838	0.72	0.914	0.831	**1**	
Kielmeyera coriacea	310	0.847	0.736	0.874	0.909	**1**	Yes
Kielmeyera grandiflora	51	0.961	0.626	0.712	**1**	0.469	Yes
Kielmeyera rosea	26	0.963	0.632	0.875	**1**	0.479	Yes
Kielmeyera rubriflora	271	0.865	0.659	0.849	0.892	**1**	Yes
Kielmeyera speciosa	59	0.964	0.659	0.447	**1**	0.209	Yes
Licania dealbata	98	0.946	0.71	0.853	**1**	0.909	Yes
Licania humilis	123	0.912	0.719	0.85	**1**	0.953	Yes
Luetzelburgia auriculata	153	0.918	0.656	0.89	0.605	**1**	
Lychnophora ericoides	184	0.915	0.533	0.524	**1**	0.271	Yes
Machaerium opacum	162	0.913	0.661	0.858	**1**	0.795	Yes
Maprounea brasiliensis	92	0.941	0.541	0.907	**1**	0.784	Yes
Mezilaurus crassiramea	31	0.935	**1**	0.525	0.14	0.274	
Mimosa claussenii	195	0.912	0.72	0.551	**1**	0.33	Yes
Mimosa decorticans	13	0.985	0.363	0.198	**1**	0.195	Yes
Mimosa interrupta	21	0.978	0.688	**1**	0.044	0.895	
Mimosa obovata	53	0.943	0.623	0.957	0.744	**1**	
Mimosa pithecolobioides	50	0.964	0.592	0.786	**1**	0.747	Yes
Mimosa pteridifolia	87	0.941	0.78	**1**	0.803	0.913	
Mimosa sericantha	51	0.967	0.743	0.945	0.782	**1**	
Mimosa verrucosa	116	0.936	0.536	0.645	0.531	**1**	
Oxandra reticulata	48	0.937	0.732	0.69	0.557	**1**	
Paralychnophora bicolor	65	0.957	0.575	0.949	**1**	0.684	Yes
Peltogyne confertiflora	130	0.903	0.755	0.873	0.558	**1**	
Persea venosa	61	0.946	0.516	0.628	**1**	0.597	Yes
Piptocarpha rotundifolia	281	0.865	0.714	0.929	0.947	**1**	Yes

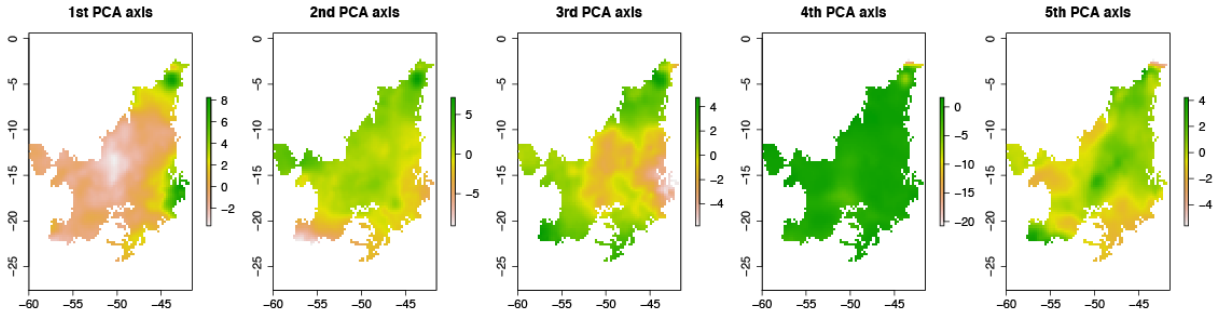
Pithecellobium gummiferum	127	0.922	0.761	0.753	**1**	0.528	Yes
Plenckia populnea	247	0.881	0.625	0.825	**1**	0.749	Yes
Rauvolfia weddeliana	52	0.967	0.367	**1**	0.357	0.709	
Rourea induta	271	0.863	0.855	0.938	0.869	**1**	Yes
Salacia crassifolia	57	0.959	0.888	0.991	**1**	0.503	Yes
Salacia grandifolia	30	0.974	0.323	0.758	**1**	0.606	Yes
Sclerolobium aureum	185	0.873	0.727	0.869	0.866	**1**	
Stryphnodendron adstringens	261	0.865	0.67	0.862	0.936	**1**	Yes
Stryphnodendron rotundifolium	143	0.893	0.797	0.828	0.951	**1**	Yes
Syagrus comosa	114	0.917	0.852	0.921	0.931	**1**	Yes
Zeyheria montana	294	0.861	0.769	0.869	0.991	**1**	Yes



Appendix 5: Interpolated layers of variance for each bioclimatic variable analyzed in stability tests. Colors of each cell represent the variance of the environmental layer, estimated by comparing cell values in present, 6,000 ybp, 21,000 ybp and 120,000 ybp scenarios. All layers were derived from WorldClim bioclimatic layers (Hijmans *et al*. 2005). (continue)



Appendix 5 (continued): Interpolated layers of variance for each bioclimatic variable analyzed in stability tests. Colors of each cell represent the variance of the environmental layer, estimated by comparing cell values in present, 6,000 ybp, 21,000 ybp and 120,000 ybp scenarios. All layers were derived from WorldClim bioclimatic layers (Hijmans *et al*. 2005).



Appendix 6: Interpolated raster layers of the five most informative (cumulative importance > 0.75) axis of Principal Component Analysis performed for the variances of each bioclimatic variable analyzed in stability tests.

5. Considerações Finais

Com o desenvolvimento de marcadores microssatélites específicos para o grupo, como o apresentado no capítulo 2, foi possível descrever a diversidade genética das espécies do grupo e posteriormente aplicar estas informações a estudos históricos, taxonômicos e de ecologia reprodutiva. Nós também pudemos concluir que ambas as espécies são diplóides, uma informação importante para estudos genéticos e reprodutivos.

O capítulo 3, por sua vez, se propôs a caracterizar a diversidade genética do grupo e discutir os padrões filogeográficos das espécies *Rauvolfia weddelliana* e *Rauvolfia gracilis* dentro do contexto das mudanças climáticas do Quaternário. Neste capítulo, pudemos concluir que as linhagens deste grupo se diversificaram inicialmente no Mioceno, com padrões demográficos de refúgios indicando refúgios interglaciais durante o médio Holoceno. Nossos resultados também indicaram inconsistências taxonômicas na atual divisão das espécies, com uma linhagem críptica ao sul de *Rauvolfia weddelliana* sendo mais geneticamente diferenciada do que *Rauvolfia gracilis*.

Por fim, no capítulo 3, nós testamos o alcance dos modelos discutidos no capítulo 2 em comparação a outros processos biogeográficos, e encontramos correlação entre o endemismo de espécies de plantas no Cerrado e sua elevação e marginalidade das distribuições. Nós também reconstruímos modelos de vegetação e concluímos que a maioria das espécies savânicas, presentes nas chapadas, expandiram sua distribuição durante o glacial e coincidem com o modelo de refúgios interglaciais.